Collapsing the Hierarchy of Compressed Data Structures: Suffix Arrays in Optimal Compressed Space

Dominik Kempa*
Department of Computer Science,
Stony Brook University,
Stony Brook, NY, USA
kempa@cs.stonybrook.edu

Tomasz Kociumaka
Max Planck Institute for Informatics,
Saarland Informatics Campus,
Saarbrücken, Germany
tomasz.kociumaka@mpi-inf.mpg.de

Abstract

The last two decades have witnessed a dramatic increase in the amount of highly repetitive datasets consisting of sequential data (strings, texts). Processing these massive amounts of data using conventional data structures is infeasible. This fueled the development of compressed text indexes, which efficiently answer various queries on a given text, typically in polylogarithmic time, while occupying space proportional to the compressed representation of the text. There exist numerous structures supporting queries ranging from simple "local" queries, such as random access, through more complex ones, including longest common extension (LCE) queries, to the most powerful queries, such as the suffix array (SA) functionality. Alongside the rich repertoire of queries followed a detailed study of the trade-off between the size and functionality of compressed indexes (see: Navarro; ACM Comput. Surv. 2021). It is widely accepted that this hierarchy of structures tells a simple story: the more powerful the queries, the more space is needed. On the one hand, random access, the most basic query, can be supported using $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$ space (where n is the length of the text, σ is the alphabet size, and δ is the text's substring complexity), which is known to be the asymptotically smallest space sufficient to represent any string with parameters n, σ , and δ , (Kociumaka, Navarro, and Prezza; IEEE Trans. Inf. Theory 2023). The other end of the hierarchy is occupied by indexes supporting the suffix array queries. The currently smallest one takes $\mathcal{O}(r\log\frac{n}{r})$ space, where $r\geq\delta$ is the number of runs in the Burrows-Wheeler Transform of the text (Gagie, Navarro, and Prezza; J. ACM 2020).

We present a new compressed index, referred to as δ -SA, that supports the powerful SA functionality and needs only $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$ space. This collapses the hierarchy of compressed data structures into a single point: The space required to represent the text is simultaneously sufficient to efficiently support the full SA functionality. Since suffix array queries are the most widely utilized queries in string processing and data compression, our result immediately improves the space complexity of dozens of algorithms, which can now be executed in δ -optimal compressed space. The δ -SA supports both suffix array and inverse suffix array queries in $\mathcal{O}(\log^{4+\epsilon} n)$ time (where $\epsilon > 0$ is any predefined constant).

Our second main result is an $\mathcal{O}(\delta \text{ polylog } n)$ time construction of the δ -SA from the Lempel–Ziv (LZ77) parsing of the text. This is the first algorithm that builds an SA index in *compressed time*, i.e., time nearly linear in the compressed input size. For highly repetitive texts, this is up to exponentially faster than the previously best algorithm, which builds an $\mathcal{O}(r \log \frac{n}{r})$ -size index in $\mathcal{O}(\sqrt{\delta n} \operatorname{polylog} n)$ time.

To obtain our results, we develop numerous new techniques of independent interest. This includes deterministic restricted recompression, δ -compressed string synchronizing sets, and their construction in compressed time. We also improve many other auxiliary data structures; e.g., we show the first $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$ -size index for LCE queries along with its efficient construction from the LZ77 parsing.

^{*}Supported by the Simons Foundation Junior Faculty Fellowship.

1 Introduction

The last few decades witnessed explosive growth in the amount of data humanity generates and needs to process. Many rapidly expanding datasets consist of sequential (textual) data, such as source code in version control systems [Nav21b], results of web crawls [FM10], versioned document collections (such as Wikipedia) [KK20a], and, perhaps most notably, biological sequences [PHR00, BDY16]. The sizes of these datasets already reach petabytes [HPWO19] and are predicted to still get orders of magnitudes larger [Nat22]. One of the key characteristics of this data, and what turns searching such datasets into the ultimate needle-in-a-haystack scenario, is that none of it can be discarded: in computational biology, the presence or lack of disease can depend on a single mutation [PHR00, Nat22], whereas in source code repositories, a bug could be a result of a single typo.

What comes to the rescue is that these datasets are extremely redundant, e.g., genomic databases are known to be up to 99.9% repetitive [PHR00]. Researchers have therefore turned their attention to techniques from lossless data compression. Compressing alone is not enough, however, as this renders the text unreadable. The solution lies in incorporating techniques from data compression directly into the design of *compressed algorithms* and *compressed data structures*:

- To date, compressed algorithms have been developed for numerous problems, ranging from exact [Gaw11, Gaw13, Jeż15, ABBK17, GG22] and approximate string matching [ABBK17, BKW19, CKW20], via computing edit distance [HLLW13, Tis15, ABBK17, GKLS22], to fundamental linear algebra operations (such as inner product, matrix-vector multiplication, and matrix multiplication) ubiquitous in machine learning [ABBK20, FMG⁺22, FGK⁺22].
- Same can be said about data structures. One can keep the data in compressed form and, at the price of a moderate (typically polylogarithmic) increase in space complexity, efficiently support various queries on the original (uncompressed) text. The currently supported queries range from the fundamental local queries like random access [BLR+15, GJL21, BCG+21], through less local rank and select [PNB17, Pre19, BCG+21] or longest common extension (LCE) queries [NII+16, I17, GKK+18], to the most powerful and complex queries like pattern matching [CN11, GGK+12, GGK+14, CNP21, DNP21, CEK+21, KNO22] and full suffix array functionality [GNP20]. The suffix array queries that, given a rank i ∈ [1..n], ask for the starting position of the ith lexicographically smallest suffix of the length-n text, are known to be particularly powerful, as they form the backbone of dozens of string processing and data compression algorithms [Gus97, ABM08].

As the field matured, numerous ways to classify and compare different compressed structures emerged [KP18, Nav21a, KNP23], resulting in hierarchies of structures ordered according to their size and functionality. As expected, structures supporting the most basic queries (such as random access) occupy the low-space regime [KNP23], while the most powerful indexes supporting suffix array functionality, such as [GNP20], require the most space. The natural question was thus raised: How much space is required to efficiently support each functionality? Kociumaka, Navarro, and Prezza [KNP23] recently proved that, letting δ be the substring complexity of the text, n be its length, and σ be the size of the alphabet, a text can be represented in $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$ space, and this bound is asymptotically tight as a function of δ , n, and σ . Simultaneously, they showed that it is possible to support random access and pattern-matching queries in the same space (see also [KNO22] for improvements of pattern matching query time). Given this situation, we thus ask:

What is the space required to efficiently support the most powerful queries, such as the suffix array functionality?

Our Results In this paper, we establish the following two main results:

1. We develop the first data structure, referred to as δ -SA, that uses only $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$ space and supports efficient suffix array queries (more precisely, it answers SA and inverse SA queries in $\mathcal{O}(\log^{4+\epsilon} n)$ time, where $\epsilon > 0$ is any given constant). This collapses the existing rich hierarchy of compressed data structures (see [KP18, GNP20, KK20a, Nav21a, KNP23]) into a single point: In view of our result, $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$ is the fundamental space complexity for compressed text indexing since, on the one hand,

- such space is required to represent the string [KNP23] (and this bound holds for all combinations of n, σ , and δ) and, on the other hand, it is already sufficient to support the powerful SA queries. Since the suffix array queries constitute the fundamental building block of string processing algorithms [Gus97, ABM08], our result immediately implies that dozens of algorithms can be executed in this δ -optimal space.
- 2. We present an algorithm that constructs the δ -SA in $\mathcal{O}(\delta \text{ polylog } n)$ time from the Lempel–Ziv (LZ77) parsing of text [ZL77]. This is the first construction of an SA index running in compressed time, i.e., in time nearly-linear in the compressed input size. The relevance of this result lies in the fact that LZ77 can be very efficiently approximated (using, e.g., [KVNP20]) and then converted (in compressed time) into the canonical greedy form (see Proposition 2.3). At the same time, LZ77 is already strong enough to compress any string into the δ -optimal size $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$ [KNP23]. This makes LZ77 the perfect input in any pipeline of algorithms running in compressed time. The only similar prior construction is the $\mathcal{O}(\delta \text{ polylog } n)$ -time construction of run-length-encoded Burrows-Wheeler Transform (BWT) from the LZ77 parsing [KK20a]. The similarity lies in the fact that RLBWT is one of the components of the r-index of Gagie, Navarro, and Prezza [GNP20], which is also capable of answering SA queries. Our construction, however, is much stronger than [KK20a]:
 - Our algorithm builds a fully-functional SA index (the δ -SA), whereas the construction from [KK20a] builds only the run-length-encoded BWT [BW94], which is just a single component of the index of [GNP20]. To date, the fastest algorithm building the complete index of [GNP20] based on the run-length-encoded BWT required $\mathcal{O}(\sqrt{rn} \text{ polylog } n) = \mathcal{O}(\sqrt{\delta n} \text{ polylog } n)$ time [GT23].
 - The δ -SA uses the δ -optimal space, while the index of [GNP20] uses more space, i.e., $\mathcal{O}(r \log \frac{n}{r})$, where $r \geq \delta$ is the number of runs in the BWT [BW94].
 - Our algorithm is deterministic, whereas [KK20a] only provides a Las-Vegas randomized procedure.

On the way to our main results, we also achieve several auxiliary goals of independent interest. In particular, we describe the first data structure efficiently answering longest common extension (LCE) queries using the δ -optimal space of $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$. Moreover, we show how to deterministically construct it from the LZ77 parsing in $\mathcal{O}(\delta \operatorname{polylog} n)$ time (Theorem 5.25). We also obtain the first analogous construction of a data structure that supports random-access queries in $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$ space (Theorem 5.24); the previous such indexes [KNP23, KNO22] only had $\Omega(n)$ -time randomized construction algorithms.

One of the biggest technical hurdles to obtaining the above results is to simultaneously achieve

- (a) δ -optimal space,
- (b) polylogarithmic worst-case query time, and
- (c) construction in compressed time (preferably deterministic)

for every component of the structure. Satisfying any two out of three would already constitute an improvement compared to the state-of-the-art SA indexes and their construction algorithms [GNP20, KK20a]. We nevertheless show that simultaneously satisfying all three is possible. Our main new techniques to achieve this are: (1) deterministic restricted recompression, and (2) δ -compressed string synchronizing sets. Restricted recompression is a technique proposed in [KRRW23] that, as shown in [KNP23], allows constructing an RLSLP (i.e., a run-length grammar; see Section 5.1) representing the text in δ -optimal space $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$. The analysis in [KNP23], however, is probabilistic, and hence yields only a Las-Vegas randomized algorithm that takes $\mathcal{O}(n)$ expected time. Here (Section 5.2), we improve it not only by proposing an explicit construction, resulting in the first $\mathcal{O}(n)$ -time deterministic algorithm, but we also show how to achieve $\mathcal{O}(\delta \text{ polylog } n)$ -time construction from the LZ77 parsing. The second main new technical contribution is developing δ -compressed string synchronizing sets. String synchronizing sets [KK19] are a powerful symmetry-breaking mechanism with numerous applications, including algorithms for longest common substrings computation [CKPR21], indexing packed strings [KK19, DFH⁺20, KK23], dynamic suffix array [KK22], converting between compressed representations [KK20a], and quantum string algorithms [AJ22, JN23]. For a given parameter $\tau \in [1..n]$. this technique selects $\mathcal{O}(n/\tau)$ synchronizing positions so that positions with matching contexts are treated consistently and (except for highly periodic regions of the text) there is at least one synchronizing position among any τ consecutive positions. In [KK20a], LZ77-compressed synchronizing sets (i.e., synchronizing sets represented by synchronizing positions located close to LZ77 phrase boundaries [ZL77]) were used to obtain an $\mathcal{O}(\delta \text{ polylog } n)$ -time algorithm for converting the LZ77 parsing into the run-length compressed BWT [BW94]. Their technique, however, cannot be utilized here due to three major obstacles: First, [KK20a] uses $\Omega(z \log n)$ space (where $z \geq \delta$ is the size of the LZ77 parsing), and hence does not meet the δ -optimal space bound requirement. Second, the algorithm in [KK20a] is able to infer some suffix ordering only for a batch of suffixes. In other words, it is an offline solution to the problem stated in this paper. Obtaining an online solution (i.e., a data structure) requires a different approach. Finally, the synchronizing set construction used in [KK20a] is Las-Vegas randomized and hence does not satisfy our goal of achieving deterministic construction. To overcome the first obstacle, rather than storing the synchronizing positions around the LZ77 phrase boundaries, we store them in what we call a cover: the set of positions in the text covered by the leftmost occurrences of substrings of some fixed length (see Section 4). This lets us bound the number of stored synchronizing positions in terms of the substring complexity δ (see Section 5.4.3). The bulk of our paper is devoted to overcoming the second obstacle. We show how to combine weighted range counting and selection queries [Cha88] with the "range refinement" technique inspired by dynamic suffix arrays [KK22], which gradually shrinks the range of SA to contain only suffixes prefixed with a desired length- 2^k string, to obtain the SA functionality. This requires substantial modifications compared to [KK22] since dynamic suffix arrays are not compressed (they use $\mathcal{O}(n \text{ polylog } n)$ space). Finally, to address the third challenge, we utilize a novel construction of synchronizing sets using restricted recompression, binding our last problem to the first technique. This is similar to [KK22], except that here we avoid the $\mathcal{O}(\log^* n)$ space increase since our structure is static. Instead, we carefully design a potential function that guides the recompression algorithm so that the outcome is deterministically as good as it would have been in expectation if we used randomization. We give a more detailed overview of our techniques in Section 3. Our final result is summarized as follows.

Theorem 1.1 (δ -SA). Given the LZ77 parsing of $T \in [0...\sigma)^n$ and any constant $\epsilon \in (0,1)$, we can in $\mathcal{O}(\delta \log^7 n)$ time construct a data structure of size $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$ (where δ is the substring complexity of T) that, given any position $i \in [1..n]$, returns the values SA[i] and $SA^{-1}[i]$ in $\mathcal{O}(\log^{4+\epsilon} n)$ time. The construction algorithm is deterministic, and the running times are worst-case.

Related Work In this paper, we focus on algorithms and data structures working for highly repetitive strings T, which can be defined as those for which either of the values: z(T) (the size of the LZ77 parsing [ZL77]), $\gamma^*(T)$ (the size of the smallest string attractor [KP18]), $g^*(T)$ (the size of the smallest context-free grammar [KY00, Ryt03, $CLL^{+}05$]), r(T) (the number of runs in the BWT [BW94]), or $\delta(T)$ (the substring complexity [KNP23]) are significantly smaller than n (the list of such measures goes on [SS82, KMS⁺03, KN10]; see [Nav21a] for a survey). We can use either of them, since a series of papers [Ryt03, CLL+05, KP18, GNP18, KK20a, KS22, KNP23] demonstrates that the ratio between any two of these values is $\mathcal{O}(\text{polylog }n)$ for every text T of length n. The redundancy captured by these measures is present in modern massive datasets. A lot of the earlier work on small-space data structures, however, focused on reducing the sizes of structures relative to the size of text, i.e., $\mathcal{O}(n\log\sigma)$ bits (assuming $T\in[0...\sigma)^n$) or, a step further, on achieving some variant of the kth order entropy bound $\mathcal{O}(nH_k(T)) + o(n\log\sigma)$, i.e., removing the "statistical" redundancy caused by skewed frequencies of individual symbols or short substrings of length $o(\log_{\sigma} n)$. Some of the most popular structures in this setting include those answering rank and select queries, such as wavelet trees [GGV03], or those with pattern matching and suffix array or suffix tree functionality, such as the FM-index [FM05], the compressed suffix array (CSA) [GV05], or the compressed suffix tree (CST) [Sad02]. Many of these structures have subsequently been implemented and are now available via libraries, such as sds1 [GBMP14]. The construction of these indexes has also received a lot of attention and nowadays, most of them can be constructed very efficiently [HSS09, Bel14, MNN17, Kem19]. Recently, new $\mathcal{O}(n \log \sigma)$ -bit indexes with CSA and CST capabilities have been proposed that also admit o(n)-time construction [KK23] if $\log \sigma = o(\sqrt{\log n})$. We refer to [NM07, Nav14, BN14, Nav16] for further details.

Organization of the Paper After introducing the basic notation and tools in Section 2, we present a technical overview in Section 3. In Section 4, we then describe the notion of string covers, which we

 $^{^{1}}$ We did not aim to optimize the polylog n factor in the construction algorithm. In particular, we utilized existing procedures whose running time has not been optimized either.

subsequently combine with the new deterministic restricted recompression in Section 5. Next, we describe the auxiliary structures supporting range queries on various grids (Section 6) and the so-called modular constraint queries (Section 7). Finally, in the main Section 8, we present our new index, the δ -SA, which answers SA and inverse SA queries.

2 Preliminaries

Basic definitions A *string* is a finite sequence of characters from a given alphabet Σ . The length of a string S is denoted |S|. For $i \in [1...|S|]^2$, the ith character of S is denoted S[i]. A substring of S is a string of the form $S[i..j) = S[i]S[i+1] \cdots S[j-1]$ for some $1 \le i \le j \le |S| + 1$. Substrings the form S[1...j) and S[i...|S|] are called *prefixes* and *suffixes*, respectively. We use \overline{S} to denote the reverse of S, i.e., $S[|S|] \cdots S[2]S[1]$. We denote the concatenation of two strings U and V, that is, $U[1] \cdots U[|U|]V[1] \cdots V[|V|]$, by UV or $U \cdot V$. Furthermore, $S^k = \bigoplus_{i=1}^k S$ is the concatenation of $k \in \mathbb{Z}_{\geq 0}$ copies of S; note that $S^0 = \varepsilon$ is the *empty string*. A nonempty string S is said to be *primitive* if it cannot be written as $S = U^k$, where $k \ge 2$. An integer $p \in [1..|S|]$ is a period of S if S[i] = S[i+p] holds for every $i \in [1..|S|-p]$. We denote the shortest period of S as per(S). For every $S \in \Sigma^+$, we define the infinite power S^{∞} so that $S^{\infty}[i] = S[1 + (i - 1) \mod |S|]$ for $i \in \mathbb{Z}$. In particular, $S = S^{\infty}[1 ... |S|]$. The rotation operation $\operatorname{rot}(\cdot)$, given a string $S \in \Sigma^+$, moves the last character of S to the front so that $rot(S) = S[|S|] \cdot S[1...|S|-1]$. The inverse operation $rot^{-1}(\cdot)$ moves the first character of S to the back so that $rot^{-1}(S) = S[2..|S|] \cdot S[1]$. For an integer $s \in \mathbb{Z}$, the operation $\text{rot}^s(\cdot)$ denotes the |s|-time composition of rot(·) (if $s \ge 0$) or rot⁻¹(·) (if $s \le 0$). Strings S, S'

```
SA[i]
        T[SA[i] \dots n]
    19
1
2
    14
         aababa
    5
         aababababaababa
    17
    12
         abaababa
    3
         abaabababaababa
    15
         ababa
    10
         ababaababa
9
    8
         abababaababa
10
    6
         ababababaababa
11
    18
         ba
12
    13
         baababa
    4
13
         baababababaababa
    16
14
         baba
15
    11
         babaababa
16
    2
         babaabababababa
    9
17
         bababaababa
18
         babababaababa
```

suffix array.

are cyclically equivalent if $S' = \operatorname{rot}^s(S)$ for some $s \in \mathbb{Z}$. By $\operatorname{lcp}(U, V)$ (resp. $\operatorname{lcs}(U, V)$) we denote the length of the longest common prefix (resp. suffix) of U and V. For any string $S \in \Sigma^*$ and any $j_1, j_2 \in [1 ... |S|]$, we denote $\operatorname{LCE}_S(j_1, j_2) = \operatorname{lcp}(S[j_1 ... |S|], S[j_2 ... |S|])$.

We use \leq to denote the order on Σ , extended to the *lexicographic* order on Σ^* so that $U, V \in \Sigma^*$ satisfy $U \leq V$ if and only if either (a) U is a prefix of V, or (b) U[1..i) = V[1..i) and $U[i] \prec V[i]$ holds for some $i \in [1..\min(|U|, |V|)]$.

Definition 2.1. For any $T \in \Sigma^n$, $P \in \Sigma^*$, and integer $\ell \geq 0$, we define

```
\begin{split} \operatorname{Occ}_{\ell}(P,T) &= \{j' \in [1 \ldots n] : \operatorname{lcp}(P,T[j' \ldots n]) \geq \min(|P|,\ell)\}, \\ \operatorname{RangeBeg}_{\ell}(P,T) &= |\{j' \in [1 \ldots n] : T[j' \ldots n] \prec P \text{ and } j' \notin \operatorname{Occ}_{\ell}(P,T)\}|, \\ \operatorname{RangeEnd}_{\ell}(P,T) &= \operatorname{RangeBeg}_{\ell}(P,T) + |\operatorname{Occ}_{\ell}(P,T)|. \end{split}
```

When $\ell = |P|$, we simply write $\operatorname{Occ}(P,T)$, Range $\operatorname{Beg}(P,T)$, and Range $\operatorname{End}(P,T)$. Moreover, by $\operatorname{Occ}_{\ell}(j,T)$, Range $\operatorname{Beg}_{\ell}(j,T)$, and Range $\operatorname{End}_{\ell}(j,T)$ we mean, respectively, $\operatorname{Occ}_{\ell}(T[j \dots n],T)$, Range $\operatorname{Beg}_{\ell}(T[j \dots n],T)$, and Range $\operatorname{End}_{\ell}(T[j \dots n],T)$.

Remark 2.2. Note that the above generalizes for the notation for ranges used in [KK23] (where the parameter ℓ was not used) and from [KK22] (where it was only defined for patterns of the form P = T[j ...n]). Here, we obtain all these notations as special cases.

Suffix array For any $T \in \Sigma^n$ (where $n \ge 1$), the *suffix array* $SA_T[1 ... n]$ of T is a permutation of [1 ... n] such that $T[SA_T[1] ... n] \prec T[SA_T[2] ... n] \prec \cdots \prec T[SA_T[n] ... n]$, i.e., $SA_T[i]$ is the starting position of the

 $^{^2 \}text{For } i,j \in \mathbb{Z} \text{, denote } [i\mathinner{\ldotp\ldotp} j] = \{k \in \mathbb{Z} : i \leq k \leq j\}, \ [i\mathinner{\ldotp\ldotp} j) = \{k \in \mathbb{Z} : i \leq k < j\}, \ \text{and} \ (i\mathinner{\ldotp\ldotp} j] = \{k \in \mathbb{Z} : i < k \leq j\}.$

lexicographically ith suffix of T; see Fig. 1 for an example. The inverse suffix array $\mathrm{ISA}_T[1\ldots n]$ (also denoted $\mathrm{SA}_T^{-1}[1\ldots n]$) is the inverse permutation of SA_T , i.e., $\mathrm{ISA}_T[j]=i$ holds if and only if $\mathrm{SA}_T[i]=j$. Intuitively, $\mathrm{ISA}_T[j]$ stores the lexicographic rank of $T[j\ldots n]$ among the suffixes of T. Note that if $T\neq \varepsilon$, then $\mathrm{Occ}_\ell(P,T)=\{\mathrm{SA}_T[i]:i\in(\mathrm{RangeBeg}_\ell(P,T)\ldots\mathrm{RangeEnd}_\ell(P,T)]\}$ holds for every $P\in\Sigma^*$ and $\ell\geq 0$. Whenever T is clear from the context, we drop the subscript in SA_T and ISA_T .

Substring complexity For a string $T \in \Sigma^n$ and $\ell \in \mathbb{Z}_{>0}$, we denote the number of length- ℓ substrings by $\mathrm{d}_{\ell}(T) = |\{T[i\mathinner{.\,.} i + \ell) : i \in [1\mathinner{.\,.} n - \ell + 1]\}|$; note that $\mathrm{d}_{\ell}(T) = 0$ if $\ell > n$. The substring complexity of T is defined as $\delta(T) = \max_{\ell=1}^n \frac{1}{\ell} \mathrm{d}_{\ell}(T)$ [KNP23]. On the one hand, as shown in [KNP23], the measure δ is an asymptotic lower bound for nearly all known compression algorithms and repetitiveness measures, including LZ77 [ZL77], run-length-compressed BWT [BW94], grammar compression [CLL^+05], and string attractors [KP18]. On the other hand, [KNP23] also shows that, if $\Sigma = [0\mathinner{.\,.} \sigma)$, then it is possible to represent T using $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ space (or more precisely, $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n}\log n)$ bits), and this bound is asymptotically tight as a function of n, σ , and $\delta(T)$. In other words, there is no combination of values n, σ , and $\delta(T)$ such that every string $T \in [0\mathinner{.\,.} \sigma)^n$ can be encoded using $o(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n}\log n)$ bits.

Lempel–Ziv compression A fragment $T[i...i+\ell)$ of T is a previous factor if it has an earlier occurrence in T, i.e., $\mathrm{LCE}_T(i,i') \geq \ell$ holds for some $i' \in [1...i)$. An LZ77-like factorization of T is a factorization $T = F_1 \cdots F_f$ into non-empty phrases such that each phrase F_j with $|F_j| > 1$ is a previous factor. In the underlying LZ77-like representation, every phrase $F_j = T[i...i+\ell)$ that is a previous factor is encoded as (i',ℓ) , where $i' \in [1...i)$ satisfies $\mathrm{LCE}_T(i,i') \geq \ell$ (and is chosen arbitrarily in case of multiple possibilities); if $F_j = T[i]$ is not a previous factor, we encode it as (T[i], 0).

The LZ77 factorization [ZL77] (or the LZ77 parsing) of a string T is then just an LZ77-like factorization constructed by greedily parsing T from left to right into the longest possible phrases. More precisely, the jth phrase F_j is the longest previous factor starting at position $1 + |F_1 \cdots F_{j-1}|$; if no previous factor starts there, then F_j consists of a single character. This greedy construction yields the smallest LZ77-like factorization of T [LZ76, Theorem 1]. We denote the number of phrases in the LZ77 parsing of T by z(T). For example, the text T = bbabaababababababababa of Fig. 1 has LZ77 factorization $b \cdot b \cdot a \cdot ba \cdot aba \cdot bababa \cdot ababa$ with z(T) = 7 phrases, and its LZ77 representation is (b, 0), (1, 1), (a, 0), (2, 2), (3, 3), (7, 6), (10, 5).

Proposition 2.3. Given a string T of length n, represented using an LZ77-like parsing consisting of f phrases, the LZ77 parsing of T can be constructed in $\mathcal{O}(f \log^4 n)$ time.

Proof. We use [KK20b, Theorem 6.11] to build a data structure that, for any pattern P represented by its arbitrary occurrence in T, returns the leftmost occurrence of P in T. Then, we process T from left to right constructing the LZ77 parsing of T. Suppose that we have already parsed a prefix T[1..i). We binary search for the maximum length ℓ such that the leftmost occurrence of $T[i..i+\ell)$ is $T[i'..i'+\ell)$ for some $i' \in [1..i)$. By definition of the LZ77 parsing, the next phrase is either T[i] (if $\ell = 0$) or $T[i..i+\ell)$ (otherwise). The construction time of [KK20b, Theorem 6.11] is $\mathcal{O}(f \log^4 n)$, whereas the query time is $\mathcal{O}(\log^3 n)$. For each phrase of the LZ77 parsing, we make $\mathcal{O}(\log n)$ queries, which take $\mathcal{O}(\log^4 n)$ time in total. Since $z(T) \leq f$, the overall running time is $(f \log^4 n)$.

String Synchronizing Sets

Definition 2.4 (τ -synchronizing set [KK19]). Let $T \in \Sigma^n$ be a string and let $\tau \in [1..\lfloor \frac{n}{2}\rfloor]$ be a parameter. A set $S \subseteq [1..n - 2\tau + 1]$ is called a τ -synchronizing set of T if it satisfies the following consistency and density conditions:

- 1. If $T[i ... i + 2\tau) = T[j ... j + 2\tau)$, then $i \in S$ holds if and only if $j \in S$ (for $i, j \in [1 ... n 2\tau + 1]$),
- 2. $S \cap [i ... i + \tau) = \emptyset$ if and only if $i \in R(\tau, T)$ (for $i \in [1 ... n 3\tau + 2]$), where

$$\mathsf{R}(\tau,T) := \{ i \in [1 \mathinner{.\,.} n - 3\tau + 2] : \mathrm{per}(T[i \mathinner{.\,.} i + 3\tau - 2]) \leq \tfrac{1}{3}\tau \}.$$

Remark 2.5. In most applications, we want to minimize |S|. Note, however, that the density condition imposes a lower bound $|S| = \Omega(\frac{n}{\tau})$ for strings of length $n \ge 3\tau - 1$ that do not contain substrings of length $3\tau - 1$ with period at most $\frac{1}{3}\tau$. Thus, we cannot hope to achieve an upper bound improving in the worst case upon the following one.

Theorem 2.6 ([KK19, Proposition 8.10]). For every string T of length n and parameter $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$, there exists a τ -synchronizing set S of size $|S| = \mathcal{O}(\frac{n}{\tau})$. Moreover, if $T \in [0 ... \sigma)^n$, where $\sigma = n^{\mathcal{O}(1)}$, such S can be deterministically constructed in $\mathcal{O}(n)$ time.

Model of computation We use the standard word RAM model of computation [Hag98] with w-bit machine words, where $w \ge \log n$, and all standard bit-wise and arithmetic operations taking $\mathcal{O}(1)$ time. Unless explicitly stated otherwise, we measure the space complexity in machine words.

3 Technical Overview

Let $T \in \Sigma^n$, where $\Sigma = [0..\sigma)$. Assume that T[n] is a symbol that does not occur in T[1..n). Moreover, let $\epsilon \in (0,1)$ be a constant. In this section, we give an overview of the δ -SA, which is a compressed text index that (a) takes $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ space, (b) answers SA and ISA queries on T in $\mathcal{O}(\log^{4+\epsilon}n)$ time, and (c) can be constructed from the LZ77 parsing of T in $\mathcal{O}(\delta(T)\log^7 n)$ time.

3.1 SA and ISA Queries

The Basic Idea First, we observe (Lemma 8.11) that the uniqueness of T[n] implies that $\mathrm{Occ}_{\ell}(j,T) = \{j' \in [1 ... n] : T^{\infty}[j' ... j' + \ell) = T^{\infty}[j ... j + \ell)\}$ holds for every $j \in [1 ... n]$ and $\ell \geq 0$ (cf. Definition 2.1). The main idea of the query algorithms is as follows:

- To calculate ISA[j] given $j \in [1..n]$, we compute the following three values for subsequent $k \in [4..\lceil \log n \rceil]$: the ranks $b = \text{RangeBeg}_{2^k}(j,T)$ and $e = \text{RangeEnd}_{2^k}(j,T)$ such that $\text{Occ}_{2^k}(j,T) = \{\text{SA}[i] : i \in (b..e]\}$ (as discussed in Section 2) as well as an arbitrary position $j' \in \text{Occ}_{2^k}(j,T)$ satisfying $j' = \min \text{Occ}_{2^{k+1}}(j',T)$. For k = 4, these values are computed from scratch; subsequently, we rely on the output of the preceding step. After completing the final step, we return $\text{ISA}[j] := \text{RangeEnd}_{\ell}(j,T)$, where $\ell = 2^{\lceil \log n \rceil} \geq n$. This is the correct answer because $\text{Occ}_{\ell}(j,T) = \text{Occ}_{n}(j,T) = \{j\}$.
- To calculate SA[i] given $i \in [1..n]$, we proceed similarly, that is, we compute, for $k \in [4..\lceil \log n \rceil]$, the ranks $b = \text{RangeBeg}_{2^k}(SA[i], T)$ and $e = \text{RangeEnd}_{2^k}(SA[i], T)$ as well as an arbitrary position $j' \in \text{Occ}_{2^k}(SA[i], T)$ satisfying $j' = \min \text{Occ}_{2^{k+1}}(j', T)$. After completing the final step, we return the position j' satisfying $j' \in \text{Occ}_{\ell}(SA[i], T)$ for $\ell = 2^{\lceil \log n \rceil} \geq n$. This is the correct answer because $\text{Occ}_{\ell}(SA[i], T) = \text{Occ}_{n}(SA[i], T) = \{SA[i]\}$. Note that the individual steps of computing SA[i] are different from those for ISA[j] since we are given the rank i rather than the position SA[i].

This basic framework is similar to [KK22]. The major difference, however, lies in the implementation of the "refinement" procedure: While [KK22] uses $\widetilde{\Theta}(n)$ space³, here we can only store $\widetilde{\mathcal{O}}(\delta(T))$ words. Since this space allowance can be up to exponentially smaller, a much more complex approach is required.

To implement the initial step in both queries, it suffices to store all length-16 substrings of T^{∞} in the lexicographic order, each augmented with the endpoints of the corresponding SA range and the position of the leftmost occurrence in $T^{\infty}[1..)$ (Section 8.5.1). Since T^{∞} contains fewer than $d_{16}(T) + 16 \le 16\delta(T) + 16 \le 32\delta(T)$ length-16 substrings, the resulting arrays need $\mathcal{O}(\delta(T))$ space. They are also easy to obtain from the LZ77 parsing: It suffices to consider all length-16 substrings overlapping phrase boundaries; for each of them, the leftmost occurrence and the number of occurrences can be determined using existing compressed text indexes [KK20b, Theorems 6.11 and 6.21] (note that we use these indexes solely within our construction procedure; they are not included in the δ -SA). The key difficulty is thus the refinement procedure.

³The $\widetilde{\mathcal{O}}(\cdot)$ notation hides factors polylogarithmic in the (uncompressed) input size n. In other words, for any function f, we have $\widetilde{\mathcal{O}}(f) = \mathcal{O}(f \text{ polylog } n)$. Similarly, $\widetilde{\Omega}(f) = \Omega(f/\text{ polylog } n)$ and $\widetilde{\Theta}(f) = \widetilde{\mathcal{O}}(f) \cap \widetilde{\Omega}(f)$.

Definition 3.1. For any $\ell \geq 1$ and $P \in \Sigma^+$, we define

```
\begin{aligned} \operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T) &= \{j' \in \operatorname{Occ}_{\ell}(P,T) : T[j' \mathinner{\ldotp\ldotp} n] \prec P \text{ and } j' \notin \operatorname{Occ}_{2\ell}(P,T) \}, \\ \operatorname{Pos}^{\operatorname{end}}_{\ell}(P,T) &= \{j' \in \operatorname{Occ}_{\ell}(P,T) : T[j' \mathinner{\ldotp\ldotp} n] \succ P \text{ and } j' \notin \operatorname{Occ}_{2\ell}(P,T) \}. \end{aligned}
```

We denote $\delta_{\ell}^{\mathrm{beg}}(P,T) := |\mathrm{Pos}_{\ell}^{\mathrm{beg}}(P,T)|$ and $\delta_{\ell}^{\mathrm{end}}(P,T) := |\mathrm{Pos}_{\ell}^{\mathrm{end}}(P,T)|$. For any $j \in [1 ... n]$, we then let $\mathrm{Pos}_{\ell}^{\mathrm{beg}}(j,T) := \mathrm{Pos}_{\ell}^{\mathrm{beg}}(T[j ... n],T)$ and $\mathrm{Pos}_{\ell}^{\mathrm{end}}(j,T) := \mathrm{Pos}_{\ell}^{\mathrm{end}}(T[j ... n],T)$. The values $\delta_{\ell}^{\mathrm{beg}}(j,T)$ and $\delta_{\ell}^{\mathrm{end}}(j,T)$ are defined analogously.

Let us fix $k \in [4 ... \lceil \log n \rceil)$ and denote $\ell = 2^k$. We observe that every P satisfies RangeBeg $_{2\ell}(P,T) = \text{RangeBeg}_{\ell}(P,T) + \delta^{\text{beg}}_{\ell}(P,T)$ and RangeEnd $_{2\ell}(P,T) = \text{RangeEnd}_{\ell}(P,T) - \delta^{\text{end}}_{\ell}(P,T)$ (Lemma 8.6). In particular, every $j \in [1 ... n]$ satisfies RangeBeg $_{2\ell}(j,T) = \text{RangeBeg}_{\ell}(j,T) + \delta^{\text{beg}}_{\ell}(j,T)$ and RangeEnd $_{2\ell}(j,T) = \text{RangeEnd}_{\ell}(j,T) - \delta^{\text{end}}_{\ell}(j,T)$. Thus, to refine the suffix array range, it suffices to compute any two values among $\delta^{\text{beg}}_{\ell}(j,T), \delta^{\text{end}}_{\ell}(j,T), |\text{Occ}_{2\ell}(j,T)|$ (during an ISA query) or among $\delta^{\text{beg}}_{\ell}(\text{SA}[i],T), \delta^{\text{end}}_{\ell}(\text{SA}[i],T), |\text{Occ}_{2\ell}(\text{SA}[i],T)|$ (during an SA query).

Denote $\tau = \lfloor \frac{\ell}{3} \rfloor$ and let R be a shorthand for $R(\tau, T) = \{i \in [1 ... n - 3\tau + 2] : \operatorname{per}(T[i ... i + 3\tau - 2]) \leq \frac{1}{3}\tau\}$. The refinement step during the computation of $\operatorname{ISA}[j]$ (resp. $\operatorname{SA}[i]$) works differently depending on whether $j \in R$ (resp. $\operatorname{SA}[i] \in R$), in which case we call j (resp. $\operatorname{SA}[i]$) periodic. Otherwise, the position is called nonperiodic. To distinguish these two cases, we store the set $R \cap C$, where C is a 14τ -cover of T, i.e., a subset of [1..n] including all positions covered by the leftmost occurrences of length- 14τ substrings of T (Definition 4.1). As $R \cap C$ might be large, we store its interval representation $\mathcal{I}(R \cap C)$, that is, we express $R \cap C$ as a union of disjoint integer intervals (Definition 4.2).

Observation 1: There exists a 14τ -cover C such that $|\mathcal{I}(\mathsf{C})| = \mathcal{O}(\delta(T))$ and $\mathcal{I}(\mathsf{C})$ admits a fast construction algorithm from the LZ77 parsing of T. Let C be a union of $(n-28\tau..n]$ as well as intervals $[x..x+28\tau)$ over all $x \in [1..n-28\tau]$ such that $x \equiv 1 \pmod{14\tau}$ and $x = \min{\mathrm{Occ}_{28\tau}(x,T)}$. The set C is a 14τ -cover since the leftmost occurrence of every length- 14τ substring of T can be extended into an interval $[x..x+28\tau)$ for x as above; see Lemma 4.7. Note that C is a subset of positions covered by the leftmost occurrences of all length- 28τ substring of T. By [KNP23], we thus have $|\mathsf{C}| \leq 84\tau\delta(T)$ (see also Lemma 4.7). On the other hand, C is a union of length- 28τ intervals. Thus, $|\mathcal{I}(\mathsf{C})| = \mathcal{O}(|\mathsf{C}|/\tau) = \mathcal{O}(\delta(T))$. To construct $\mathcal{I}(\mathsf{C})$, it suffices to check at most two positions around each LZ77 phrase boundary (Proposition 4.8). Using an index for finding leftmost occurrences [KK20b, Theorem 6.11], we can thus build $\mathcal{I}(\mathsf{C})$ in $\mathcal{O}(z(T))$ polylog n $= \mathcal{O}(\delta(T))$ polylog n time.

Observation 2: The above C satisfies $|\mathcal{I}(\mathsf{R}\cap\mathsf{C})| = \mathcal{O}(\delta(T))$ and $\mathcal{I}(\mathsf{R}\cap\mathsf{C})$ also admits fast construction. First, we observe that any two maximal blocks of consecutive positions in R are separated by a gap of size $\Omega(\tau)$ (Lemma 8.13). This implies that, using the above C, the interval representation of $\mathsf{C}\cap\mathsf{R}$ is of size $\mathcal{O}(\delta(T))$ (Lemma 8.14). Above, we noted that $\mathcal{I}(\mathsf{C})$ can be constructed from the LZ77 parsing in $\mathcal{O}(\delta(T) \text{ polylog } n)$ time. It remains to observe that, given $\mathcal{I}(\mathsf{C})$, constructing $\mathcal{I}(\mathsf{R}\cap\mathsf{C})$ reduces to computing the shortest periods via so-called 2-period queries (Lemma 8.15). Consequently, by utilizing the index for 2-period queries [KK20b, Theorem 6.7], we can construct $\mathcal{I}(\mathsf{R}\cap\mathsf{C})$ in $\mathcal{O}(\delta(T) \text{ polylog } n)$ time; see Proposition 8.18.

Observation 3: Using $\mathcal{I}(\mathsf{R}\cap\mathsf{C})$, we can efficiently check if $j\in\mathsf{R}$ (resp. $\mathsf{SA}[i]\in\mathsf{R}$). Recall that, at the beginning of the refinement step, we have some $j'\in\mathsf{Occ}_\ell(j,T)$ (resp. $j'\in\mathsf{Occ}_\ell(\mathsf{SA}[i],T)$). By $3\tau-1\leq\ell$ and the definition of R , we thus have $j\in\mathsf{R}$ (resp. $\mathsf{SA}[i]\in\mathsf{R}$) if and only if $j'\in\mathsf{R}$. Moreover, since j' satisfies $j'=\min\mathsf{Occ}_{2\ell}(j',T)$, and any 14τ -cover is also a 2ℓ -cover (Lemma 4.4), it thus follows that $j'\in\mathsf{C}$. Thus, $j'\in\mathsf{C}\cap\mathsf{R}$ if and only if $j'\in\mathsf{R}$. Consequently, to check if $j'\in\mathsf{R}$, it suffices to check j' belongs to any interval contained in $\mathcal{I}(\mathsf{R}\cap\mathsf{C})$. Provided that the intervals in $\mathcal{I}(\mathsf{R}\cap\mathsf{C})$ are ordered left-to-right, this takes $\mathcal{O}(\log|\mathcal{I}(\mathsf{R}\cap\mathsf{C})|)=\mathcal{O}(\log n)$ time (Proposition 8.17).

The above is a simplified analysis, and reaching $\mathcal{O}(\delta(T))$ space for each $k \in [4..\lceil \log n \rceil)$ is still not enough to achieve the δ -optimal bound of $\mathcal{O}(\delta(T)\log \frac{n\log \sigma}{\delta(T)\log n})$ for the entire structure. In our complete analysis, we prove a tighter upper bound: $|\mathcal{I}(\mathsf{R} \cap \mathsf{C})| = \mathcal{O}(\frac{1}{\ell}\mathsf{d}_{38\ell}(T) + 1)$; see Section 8.2.

The Nonperiodic Positions Assume $j \in [1 ... n] \setminus \mathbb{R}$ (resp. $SA[i] \in [1 ... n] \setminus \mathbb{R}$). We first focus on computing ISA[j]. Recall that we are given $b = \operatorname{RangeBeg}_{\ell}(j,T)$, $e = \operatorname{RangeEnd}_{\ell}(j,T)$, and some $j' \in \operatorname{Occ}_{\ell}(j,T)$ satisfying $j' = \min \operatorname{Occ}_{2\ell}(j',T)$ as input. The refinement step for nonperiodic positions first computes the position $j'' = \min \operatorname{Occ}_{2\ell}(j,T)$ (this condition implies $j'' = \min \operatorname{Occ}_{4\ell}(j'',T)$), and then the values $\delta_{\ell}^{\text{beg}}(j,T)$ and $|\operatorname{Occ}_{2\ell}(j,T)|$. By the above discussion, this is sufficient to infer $\operatorname{RangeBeg}_{2\ell}(j,T)$ and $\operatorname{RangeEnd}_{2\ell}(j,T)$. Let S be a τ -synchronizing set of T (Definition 2.4). Observe that, by $3\tau \leq \ell < n$, the uniqueness of T[n] in T yields $\operatorname{per}(T[n-3\tau+2 ... n]) > \frac{1}{3}\tau$. Thus, $n-3\tau+2 \notin \mathbb{R}$, and hence the density condition (Definition 2.4(2)) implies $S \cap [n-3\tau+2 ... n-2\tau+2) \neq \emptyset$. Consequently, $\max S \geq n-3\tau+2$, and, for every $p \in [1 ... n-3\tau+2]$, we can define $\operatorname{succ}_S(p) = \min \{s \in S : s \geq p\}$. If $j > n-3\tau-2$, the uniqueness of T[n] in T implies $\operatorname{Occ}_{\ell}(j,T) = \{j\}$ (Proposition 8.40). Thus, we henceforth assume $j \in [1 ... n-3\tau+2] \setminus \mathbb{R}$.

Observation 1: The set $\operatorname{Occ}_{2\ell}(j,T)$ can be characterized using S and range queries. First, note that $j \in [1 \dots n - 3\tau + 2] \setminus \mathbb{R}$ and the density condition (Definition 2.4(2)) yield $\operatorname{succ}_{S}(j) - j < \tau$. On the other hand, by $3\tau - 1 \leq 2\ell$ and the consistency condition (Definition 2.4(1)), every $p \in \operatorname{Occ}_{2\ell}(j,T)$ satisfies $\operatorname{succ}_{S}(p) - p = \operatorname{succ}_{S}(j) - j$. Thus, letting $x_1 = \overline{T^{\infty}[j \dots \operatorname{succ}_{S}(j))}$, and $y_1 = T^{\infty}[\operatorname{succ}_{S}(j) \dots j + 2\ell)$, the set $\operatorname{Occ}_{2\ell}(j,T)$ consists of all positions of the form $s - |x_1|$, where $s \in S$ and s is preceded by $\overline{x_1}$ and followed by y_1 in T^{∞} . In other words, $\operatorname{Occ}_{2\ell}(j,T) = \{s - |x_1| : s \in S \text{ and } T^{\infty}[s - |x_1| \dots s + |y_1|) = \overline{x_1} \cdot y_1\}$. If we further denote $c = \max \Sigma$, $x_2 = x_1 c^{\infty}$, and $y_2 = y_1 c^{\infty}$, we have $\operatorname{Occ}_{2\ell}(j,T) = \{s - |x_1| : s \in S, x_1 \leq \overline{T^{\infty}[s - 7\tau \dots s)} \prec x_2$, and $y_1 \leq T^{\infty}[s \dots s + 7\tau) \prec y_2\}$ due to $|x_1|, |y_1| \leq 7\tau$. Thus, letting $\mathcal{P} = \{(\overline{T^{\infty}[s - 7\tau \dots s)}, T^{\infty}[s \dots s + 7\tau), s) : s \in S\}$ be a set of labeled points, the set $\operatorname{Occ}_{2\ell}(j,T)$ shifted by $|x_1|$ forward consists of the labels of the points in the range $[x_1 \dots x_2) \times [y_1 \dots y_2)$.

By the above, computing $\min \operatorname{Occ}_{2\ell}(j,T)$ reduces to an orthogonal range minimum query, returning the minimum label of a point occurring in the rectangle $[x_1 \dots x_2) \times [y_1 \dots y_2)$ (Lemma 8.39). Implementing such queries, however, is challenging. First, the coordinates of points in \mathcal{P} are substrings of T^{∞} . Comparing them reduces to LCE and random access queries, and hence it suffices to only store labels of points in \mathcal{P} , i.e., the set S (Proposition 6.4). The smallest prior structure for LCE queries [117], however, does not match our space bound. To avoid this issue, we develop the first structure that uses δ -optimal space $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$. In addition, we describe its $\mathcal{O}(\delta(T)\operatorname{polylog} n)$ -time deterministic construction from the LZ77 parsing [ZL77] (see Theorem 5.25). We also describe the first deterministic construction of the δ -optimal-space structure for random access queries (Theorem 5.24). We elaborate more on these indexes in Section 3.2.

The challenge thus reduces to storing and querying S. The plain representation is too large since it is not possible to reduce |S| below $\Theta(\frac{n}{\tau})$ in the worst case (Remark 2.5). We thus need to store S in a compressed form. Our starting point is the technique introduced in [KK20a], which compresses S by only keeping elements of S that are within distance $\Theta(\tau)$ from LZ77 phrase boundaries. These LZ77-compressed τ -synchronizing sets, however, do not meet the δ -optimal space bound and come only with Las-Vegas randomized construction [KK20a]. To solve the space issue, we again employ a 14τ -cover C of T. Denote $S_{\text{comp}} = S \cap C$ and $\mathcal{P}_{\text{comp}} = \{(T^{\infty}[s - 7\tau ...s), T^{\infty}[s ...s + 7\tau), s) : s \in S_{\text{comp}}\}$.

Observation 2: We can compute x_1, x_2, y_1 , and y_2 using S_{comp} . Observe that, since $|x_1| + |y_1| = 2\ell$ and all strings in question occur near j, the difficulty lies in computing $|x_1|$. Recall that we are given some $j' \in \text{Occ}_{\ell}(j,T)$ satisfying $j' = \min \text{Occ}_{2\ell}(j',T)$ as input. By $3\tau - 1 \le \ell$, the consistency of S (Definition 2.4(1)) yields $|x_1| = \text{succ}_{S}(j) - j = \text{succ}_{S}(j') - j'$. On the other hand, since $j' = \min \text{Occ}_{2\ell}(j',T)$ and C is also a 2ℓ -cover (Lemma 4.4), it follows that $[j' ... j' + 2\ell) \cap [1...n] \subseteq C$. Thus $\text{succ}_{S}(j) - j < \tau$ implies $\text{succ}_{S}(j') \in S_{\text{comp}}$ (Lemma 8.25). Consequently, it suffices to store the sorted set S_{comp} . Given j', we can then quickly determine $\text{succ}_{S_{\text{comp}}}(j') - j' = \text{succ}_{S}(j') - j' = \text{succ}_{S}(j) - j = |x_1|$.

Observation 3: Orthogonal range minimum queries on $\mathcal{P}_{\text{comp}}$ and \mathcal{P} are equivalent. Let $x_1, x_2, y_1, y_2 \in \Sigma^*$ and let s_{comp} (resp. s) denote the output of the range minimum query in $[x_1 \dots x_2) \times [y_1 \dots y_2)$ on $\mathcal{P}_{\text{comp}}$ (resp. \mathcal{P}). Observe that $\mathsf{S}_{\text{comp}} \subseteq \mathsf{S}$ implies $s \leq s_{\text{comp}}$. To show the opposite inequality, let $s_{\min} = \min\{i \in [1 \dots n]: T^{\infty}[i - 7\tau \dots i + 7\tau) = T^{\infty}[s - 7\tau \dots s + 7\tau)\} \leq s$. Then, either $s_{\min} \in [1 \dots 7\tau] \cup (n - 7\tau \dots n]$ or $s_{\min} - 7\tau = \min \mathrm{Occ}_{14\tau}(s_{\min} - 7\tau, T)$. In both cases, $s_{\min} \in \mathsf{C}$. Moreover, by the consistency of S (Definition 2.4(1)), $s_{\min} \in \mathsf{S}$. Thus, $s_{\min} \in \mathsf{S}_{\text{comp}}$. Finally, $T^{\infty}[s_{\min} - 7\tau \dots s_{\min} + 7\tau) = T^{\infty}[s - 7\tau \dots s + 7\tau)$ implies that $(T^{\infty}[s_{\min} - 7\tau \dots s_{\min}), T^{\infty}[s_{\min} \dots s_{\min} + 7\tau)) \in [x_1 \dots x_2) \times [y_1 \dots y_2)$

```
if and only if (\overline{T^{\infty}[s-7\tau \ldots s)}, T^{\infty}[s \ldots s+7\tau)) \in [x_1 \ldots x_2) \times [y_1 \ldots y_2). Thus, s_{\text{comp}} \leq s_{\text{min}} \leq s.
```

By the above, it suffices to use $\mathcal{P}_{\text{comp}}$ during the computation of min $\text{Occ}_{2\ell}(j,T)$ (Lemma 8.39). The computation of $\delta_{\ell}^{\text{beg}}(j,T)$ and $|\text{Occ}_{2\ell}(j,T)|$ uses similar ideas, except range minimum queries are replaced with range counting (Lemmas 8.29 and 8.33). For this, we augment each point with a *weight* storing the frequency of the corresponding substring. The correctness of this follows by the local consistency of S (Lemma 8.23). To avoid double counting, we also need to ensure that no two points in $\mathcal{P}_{\text{comp}}$ coincide. To simultaneously still allow range minimum queries, we thus leave only points with the smallest labels; see Definition 6.3.

Let us now return to computing SA[i]. Observe that, to compute $\min Occ_{2\ell}(j,T)$, $\delta_{\ell}^{beg}(j,T)$, and $|Occ_{2\ell}(j,T)|$, we needed some occurrences of strings x_1 and y_1 satisfying $|x_1| = succ_{S}(j) - j$ and $T^{\infty}[j ... j + 2\ell) = \overline{x_1}y_1$. The input of the refinement procedure in the computation of SA[i], however, does not include any element of $Occ_{2\ell}(SA[i],T)$. Consequently, our query procedure performs an additional step that computes some position $p \in Occ_{2\ell}(SA[i],T)$. Such a position can be retrieved from \mathcal{P}_{comp} using the inverse of range counting, i.e., range selection queries (see Lemma 8.36). The rest of the query proceeds similarly to the computation of ISA[j] for j = SA[i], except we can now use p instead of j since $p \in Occ_{2\ell}(SA[i],T)$ implies that $Occ_{2\ell}(SA[i],T) = Occ_{2\ell}(p,T)$ (Lemma 8.11).

The remaining challenge is ensuring that S_{comp} is small. Unlike for the compressed version of R, where upper bounds on |C| and $|\mathcal{I}(C)|$ already impose an upper bound on $|\mathcal{I}(R \cap C)|$, the size of $S \cap C$ can be large if S is not constructed carefully. In Section 5, we develop a deterministic construction that ensures that the total size of S_{comp} across all levels $k \in [4.. \lceil \log n \rceil)$ of the data structure is $\mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$ (see Section 3.2).

The Periodic Positions Let us now assume $j \in \mathbb{R}$ (resp. $\mathrm{SA}[i] \in \mathbb{R}$). One of the key challenges, compared to previous work using the "refinement" framework [KK22] is as follows. The basic property of every $p \in \mathbb{R}$ (extending easily to blocks of such positions), dictating the rest of the query algorithm, is its type, defined either as -1 or +1, depending on whether the symbol following the periodic substring is larger or smaller than the symbol that would extend the period (see Section 8.4.1). Dealing with positions of each type is straightforward if $\tilde{\Theta}(n)$ space is available: We separately store all maximal blocks of positions in \mathbb{R} of each type [KK22]. In compressed space, however, we are much more constrained. For example, $j \in \mathbb{R}$ and $j' \in \mathrm{Occ}_{2\ell}(j,T)$ may have different types, so we cannot distinguish the type simply based on the occurrence of a periodic fragment. This requires numerous new and more general combinatorial properties, allowing separate processing of elements of $\mathrm{Occ}_{2\ell}(j,T)$ (resp. $\mathrm{Occ}_{2\ell}(\mathrm{SA}[i],T)$) depending on whether they are partially periodic (i.e., the length of their periodic prefix is less than 2ℓ) or fully periodic (otherwise); see Section 8.4.

3.2 Deterministic Restricted Recompression

Restricted recompression [KRRW23] is a general technique for constructing a run-length grammar (RLSLP) of a given text (see Section 5.1). Utilizing this technique, Kociumaka, Navarro, and Prezza [KNP23] proved that every $T \in [0..\sigma)^n$ can be represented using an RLSLP of size $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ and height $\mathcal{O}(\log n)$. They also showed that $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ is the asymptotically optimal space to represent a string, for all combinations of n, σ , and $\delta(T)$. Consequently, random access to T can be efficiently supported in the δ -optimal space. Finally, they developed an $\mathcal{O}(n)$ -expected-time Las-Vegas randomized construction of such RLSLP. At the heart of their construction is the problem of approximating the directed max-cut of graphs derived from partially compressed representations of T. In [KNP23], it is proved that a uniformly random partition at every level of the grammar is sufficient to achieve the δ -optimal total size in expectation. In this paper, we describe an explicit partitioning technique (Construction 5.7) resulting in the same bound on the size of the RLSLP (Section 5.2.1). The unique component of our construction is the use of a cover hierarchy (Section 4), allowing us to account for the effects of partitioning at the current level of the grammar on the properties of the grammar at all future levels. In addition, we develop an $\mathcal{O}(\delta(T)$ polylog n)-time deterministic construction algorithm of our RLSLP from the LZ77 parsing of T [ZL77].

Equipped with this RLSLP, it is relatively easy to answer random access and LCE queries on T in $\mathcal{O}(\log n)$ time (Theorems 5.24 and 5.25). Moreover, in $\mathcal{O}(\delta(T) \operatorname{polylog} n)$ time, we can derive from the RLSLP a sequence of string synchronizing sets such that, after pairwise intersection with the cover hierarchy guiding the RLSLP construction, we obtain their representation of total size $\mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$ (Proposition 5.30).

4 Optimal Compressed Space String Covers

Definition 4.1 (Cover). Let $T \in \Sigma^n$ and $\ell \in \mathbb{Z}_{>0}$. A set $C \subseteq [1..n]$ is called an ℓ -cover of T if $(\max(0, n - \ell)..n] \subseteq C$ and $[i..i + \ell) \subseteq C$ holds for every $i \in [1..n - \ell]$ satisfying $i = \min \operatorname{Occ}_{\ell}(i, T)$. In other words, C must contain the last $\min(\ell, n)$ positions of T, as well as all positions covered by the leftmost occurrences of length- ℓ substrings of T.

Definition 4.2 (Interval representation). The interval representation of a finite set $P \subseteq \mathbb{Z}_{>0}$ is the unique sequence $\mathcal{I}(P) = (a_i, t_i)_{i \in [1..m]}$ such that $a_1 < a_1 + t_1 < a_2 < a_2 + t_2 < \cdots < a_m < a_m + t_m$ and $P = \bigcup_{i \in [1..m]} [a_i \ldots a_i + t_i)$.

Definition 4.3 (Cover hierarchy). A *cover hierarchy* of a string $T \in \Sigma^+$ is an ascending sequence of sets $\mathsf{C} = (\mathsf{C}_\ell)_{\ell \in \mathbb{Z}_{>0}}$ such that C_ℓ is an ℓ -cover of T and $|\mathcal{I}(\mathsf{C}_\ell)| \leq \max(1, \frac{|\mathsf{C}_\ell|}{\ell})$.

Lemma 4.4. Let $T \in \Sigma^n$ and $\ell \in \mathbb{Z}_{>0}$. If C is an ℓ -cover of T, then it is also an ℓ' -cover of T for every $\ell' \in [1..\ell)$.

Proof. Let $\ell' \in [1 \dots \ell)$. First, note that $(\max(0, n - \ell') \dots n] \subseteq (\max(0, n - \ell) \dots n] \subseteq \mathbb{C}$. Next, let us consider $i \in [1 \dots n - \ell']$ such that $i = \min \operatorname{Occ}_{\ell'}(i, T)$. If $i + \ell > n$, then $[i \dots i + \ell') \subseteq (\max(0, n - \ell) \dots n] \subseteq \mathbb{C}$ follows as above. We can thus assume $i \in [1 \dots n - \ell]$. Observe that this implies $i = \min \operatorname{Occ}_{\ell}(i, T)$. Otherwise, there would exist $i' \in [1 \dots i)$ such that $T[i' \dots i' + \ell) = T[i \dots i + \ell)$. By $\ell' < \ell$, we then have $T[i' \dots i' + \ell') = T[i \dots i + \ell')$, contradicting $i = \min \operatorname{Occ}_{\ell'}(i, T)$. Thus, $i = \min \operatorname{Occ}_{\ell}(i, T)$, and hence $[i \dots i + \ell) \subseteq \mathbb{C}$. By $\ell' < \ell$, we thus obtain $[i \dots i + \ell') \subseteq \mathbb{C}$. □

Lemma 4.5. Let $T \in \Sigma^n$ and $\ell \in \mathbb{Z}_{>0}$. If C is an ℓ -cover of T, then $[i ... i + \ell) \cap [1 ... n] \subseteq C$ holds for every $i \in [1 ... n]$ satisfying $i = \min \operatorname{Occ}_{\ell}(i, T)$.

Proof. Let $i \in [1 ... n]$ be such that $i = \min \operatorname{Occ}_{\ell}(i, T)$. If $i + \ell > n$, then by Definition 4.1, we immediately obtain $[i ... i + \ell) \cap [1 ... n] \subseteq (\max(0, n - \ell) ... n] \subseteq \mathbb{C}$. Otherwise (i.e., $i \in [1 ... n - \ell]$), it holds $[i ... i + \ell) \cap [1 ... n] = [i ... i + \ell)$, and hence the claim follows by the second property in Definition 4.1.

Construction 4.6. Let $T \in \Sigma^n$ and $\ell \in \mathbb{Z}_{>0}$. Denoting $k = 2^{\lceil \log \ell \rceil}$, we define the set

$$\mathsf{C}(\ell,T) := \bigcup_{i \in \mathsf{I}} [i \mathinner{\ldotp\ldotp} i + 2k) \ \cup \ (\max(0,n-2k)\mathinner{\ldotp\ldotp} n].$$

where $I = \{i \in [1 ... n - 2k] : i \mod k = 1 \text{ and } i = \min Occ_{2k}(i, T)\}.$

Lemma 4.7. For every text $T \in \Sigma^+$, the family $(\mathsf{C}(\ell,T))_{\ell \in \mathbb{Z}_{>0}}$ forms a cover hierarchy. Moreover, each set in this hierarchy satisfies $|\mathsf{C}(\ell,T)| \leq \mathsf{d}_{8\ell}(T) + 8\ell \leq 16\ell\delta(T)$ and $|\mathcal{I}(\mathsf{C}(\ell,T))| \leq \frac{\mathsf{d}_{8\ell}(T)}{\ell} + 8 \leq 16\delta(T)$.

Proof. Denote n = |T|. Let us first assume that $\ell \leq \frac{1}{2}n$ is a power of two and prove that $\mathsf{C}(\ell,T)$ is an ℓ -cover. First, we note that $(\max(0,n-\ell)\dots n] \subseteq (\max(0,n-2k),\dots n] \subseteq \mathsf{C}(\ell,T)$. Let us consider any $i \in [1\dots n-\ell]$ such that $i = \min \mathsf{Occ}_{\ell}(i,T)$. If $i+2\ell > n$, then $[i\dots i+\ell) \subseteq (\max(0,n-2\ell)\dots n] \subseteq (\max(0,n-2k)\dots n] \subseteq \mathsf{C}(\ell,T)$. We can thus assume $i \in [1\dots n-2\ell]$. Define j as the largest integer such that $j \leq i$ and $j \mod \ell = 1$. Then, $i-j < \ell$ and hence $[i\dots i+\ell) \subseteq [j\dots j+2\ell)$. Moreover, we claim that $j = \min \mathsf{Occ}_{2\ell}(j,T)$. Otherwise, there would exist j' < j such that $T[j'\dots j'+2\ell) = T[j\dots j+2\ell)$. This would imply that i' = j' + (i-j) = i - (j-j'') < i satisfies $T[i'\dots i'+\ell) = T[i\dots i+\ell)$, contradicting $i = \min \mathsf{Occ}_{\ell}(i,T)$. Thus, $j \in I$ and $[i\dots i+\ell) \subseteq [j\dots j+2\ell) \subseteq \mathsf{C}(\ell,T)$. Hence, $\mathsf{C}(\ell,T)$ is an ℓ -cover of T.

Next, we shall bound $|\mathsf{C}(\ell,T)|$. Observe that if $i \in \mathsf{C}(\ell,T) \cap (2\ell \dots n-2\ell]$, then there is a position $j \in (i-2\ell \dots i]$ such that $j = \min \mathsf{Occ}_{2\ell}(j,T)$. Consequently, $i-2\ell = \min \mathsf{Occ}_{4\ell}(i-2\ell,T)$. Thus, $|\mathsf{C}(\ell,T)| \le \mathsf{d}_{4\ell}(T) + 4\ell \le 8\ell\delta(T)$.

Furthermore, observe that $C(\ell, T)$ is a union of length- 2ℓ intervals, and hence $|\mathcal{I}(C(\ell, T))| \leq \frac{1}{2\ell} |C(\ell, T)| \leq \frac{1}{\ell} |C(\ell, T)|$. Moreover, the sequence $C(\ell, T)$ is ascending when restricted to powers of two not exceeding $\frac{1}{2}n$: Each position $i \in C(\ell, T) \cap [1 ... n - 2\ell]$ is covered by the leftmost occurrence of a length- 2ℓ substring of T,

and thus it must be contained in any 2ℓ -cover. Furthermore, each position $i \in (n-2\ell ...n]$ is contained in $C(2\ell,T)$ by construction.

It remains to consider the case of arbitrary $\ell \in \mathbb{Z}_{>0}$. If $\ell > \frac{1}{2}n$, we have $\mathsf{C}(\ell,T) = [1\dots n]$. This set is trivially an ℓ -cover, it satisfies $|\mathcal{I}(\mathsf{C}(\ell,T))| = 1$ and $\mathsf{C}(\ell,T) = n < 2\ell$. If $\ell \leq \frac{1}{2}n$ is not a power of two, then we have $\mathsf{C}(\ell,T) := \mathsf{C}(k,T)$, where $k = 2^{\lceil \log \ell \rceil}$. This set is an ℓ -cover by Lemma 4.4. Furthermore, $|\mathsf{C}(\ell,T)| = |\mathsf{C}(k,T)| \leq \mathsf{d}_{4k}(T) + 4k \leq \mathsf{d}_{8\ell}(T) + 8\ell \leq 16\ell \cdot \delta(T)$ and $|\mathcal{I}(\mathsf{C}(\ell,T))| = |\mathcal{I}(\mathsf{C}(k,T))| \leq \max(1,\frac{1}{k}|\mathsf{C}(k,T)|) \leq \max(1,\frac{1}{k}|\mathsf{C}(k,T)|) \leq 16\delta(T)$.

Proposition 4.8. Let $T \in \Sigma^+$ and $\ell \in \mathbb{Z}_{\geq 0}$ Assume that, for every substring Q of T (specified with its starting position and the length), we can in $\mathcal{O}(t_{\text{minocc}})$ time compute $\min \operatorname{Occ}(Q,T)$. Given the LZ77 parsing of T, we can construct the interval representation $\mathcal{I}(\mathsf{C}(\ell,T))$ of the cover of Construction 4.6 in $\mathcal{O}(z(T) \cdot t_{\text{minocc}})$ time.

Proof. Since $C(\ell, T) = C(k, T)$ holds for $k = 2^{\lceil \log \ell \rceil}$, we can assume without loss of generality that ℓ is a power of two.

Let I and $C(\ell, T)$ be defined as in Lemma 4.7. Observe that to compute $\mathcal{I}(C(\ell, T))$, it suffices to enumerate the set I left-to-right. The elements of the sequence are then easily constructed.

The set I can be enumerated as follows. Let b_j , where $j \in [1 ... z(T)]$, denote the first position of the jth leftmost phrase in the input parsing. Observe that for every $i \in I$, there exists an index $j \in [1 ... z(T)]$ such that $i < b_j < i + 2\ell$, since otherwise $T[i ... i + 2\ell)$ would be entirely contained inside some phrase and hence have an earlier occurrence in T, contradicting $i = \min \operatorname{Occ}_{2\ell}(i,T)$. For such j we therefore have $i \in (b_j - 2\ell ... b_j)$. Note now that by taking into account the constraint $i \mod i = 1$ in the definition of I, we have $|(b_j - 2\ell ... b_j) \cap I| \leq 2$. Consequently, to enumerate I it suffices for every $j \in [1 ... z(T)]$ to inspect at most two candidate positions: the first at $i_1 = \ell \lceil \frac{b_j - 2\ell}{\ell} \rceil + 1$ and second at $i_2 = i_1 + \ell$ (if $i_2 < b_j$). For each candidate i, testing reduces to checking if it holds $i = \min \operatorname{Occ}_{2\ell}(i,T)$. The latter check can be implemented in $\mathcal{O}(t_{\min occ})$ time by computing $i' = \min \operatorname{Occ}(T[i ... i + 2\ell), T)$ and checking if i' = i. Over all $j \in [1 ... z(T)]$, we thus spend $\mathcal{O}(z(T) \cdot t_{\min occ})$ time. Note that some candidates i may repeat across different $j \in [1 ... z(T)]$, but it is easy to detect and eliminate such candidates by storing the last two candidate positions. \square

By [KK20b, Theorem 6.11], the queries specified in the statement of Proposition 4.8 can be answered in $\mathcal{O}(\log^3 n)$ time after $\mathcal{O}(z(T)\log^4 n)$ -time preprocessing. Thus, we get the following

Corollary 4.9. Given the LZ77 parsing of a string $T \in \Sigma^n$ and an integer $\ell \in \mathbb{Z}_{>0}$, the interval representation $\mathcal{L}(\mathsf{C}(\ell,T))$ of the cover of Construction 4.6 can be built in $\mathcal{O}(z(T) \cdot \log^4 n)$ time.

Definition 4.10 (Compressed representation). Let $T \in \Sigma^n$ and $P \subseteq [1..n]$. For every $\ell \ge 1$, we define the compressed representation of P as $\text{comp}_{\ell}(P,T) := P \cap C(\ell,T)$, where $C(\ell,T)$ is the ℓ -cover of T defined in Construction 4.6.

5 From LZ Parsing to Restricted Recompression

As shown in [KNP23], every string $T \in [0...\sigma)^n$ can be represented using a run-length context-free grammar of size $\mathcal{O}(\delta(T) \cdot \log \frac{n \log \sigma}{\delta(T) \log n})$. In this section, we revise the original construction in order to achieve a deterministic $\mathcal{O}(\delta(T) \log^r n)$ -time algorithm that builds such a grammar from the LZ77 representation of T. The underlying modifications are also crucial to make sure that the synchronizing sets defined in [KRRW23] (using a similar run-length context-free grammar) admit compressed representations of $\mathcal{O}(\delta(T) \cdot \log \frac{n \log \sigma}{\delta(T) \log n})$ size in total.

5.1 Preliminaries

For a context-free grammar \mathcal{G} , we denote by $\Sigma_{\mathcal{G}}$ and $\mathcal{N}_{\mathcal{G}}$ the set of non-terminals and the set of terminals, respectively. The set of symbols is $\mathcal{S}_{\mathcal{G}} := \Sigma_{\mathcal{G}} \cup \mathcal{N}_{\mathcal{G}}$. A straight-line grammar (SLG) is a context-free grammar \mathcal{G} such that:

• each non-terminal $A \in \mathcal{N}_{\mathcal{G}}$ has a unique production $A \to \mathsf{rhs}_{\mathcal{G}}(A)$, where $\mathsf{rhs}_{\mathcal{G}}(A) \in \mathcal{S}_{\mathcal{G}}^*$,

• the set of symbols $\mathcal{S}_{\mathcal{G}}$ admits a partial order \prec such that $B \prec A$ if B appears in $\mathsf{rhs}(A)$.

The expansion function $\exp_{\mathcal{G}}: \mathcal{S}_G \to \Sigma_{\mathcal{G}}^*$ is defined as follows:

$$\exp_{\mathcal{G}}(A) = \begin{cases} A & \text{if } A \in \Sigma_{\mathcal{G}}, \\ \exp(A_1) \exp(A_2) \cdots \exp(A_a) & \text{if } A \in \mathcal{N}_{\mathcal{G}} \text{ with } \mathsf{rhs}_{\mathcal{G}}(A) = A_1 A_2 \cdots A_a. \end{cases}$$

In particular, the expansion $\exp_{\mathcal{G}}(S)$ of a starting symbol $S \in \mathcal{S}$ is the unique string represented by \mathcal{G} . Moreover, $\exp_{\mathcal{G}}$ is lifted to $\exp_{\mathcal{G}}: \mathcal{S}_{\mathcal{G}}^* \to \Sigma_{\mathcal{G}}^*$ by setting $\exp_{\mathcal{G}}(A_1 \cdots A_a) = \exp_{\mathcal{G}}(A_1) \cdots \exp_{\mathcal{G}}(A_a)$ for $A_1 \cdots A_a \in \mathcal{S}_{\mathcal{G}}^*$. When the grammar \mathcal{G} is clear from context, we omit the superscript.

For every symbol $A \in \mathcal{S}$, we also define $\mathsf{LS}(A)$ and $\mathsf{RS}(A)$ as the leftmost and the rightmost character of $\exp(A)$, respectively, with $\mathsf{LS}(A) = \mathsf{RS}(A) = \varepsilon$ if $\exp(A) = \varepsilon$. These values can be obtained in $\mathcal{O}(|\mathcal{G}|)$ time by processing symbols of \mathcal{G} according to the underlying order \prec .

The parse tree $\mathcal{T}(A)$ of a symbol $A \in \mathcal{S}$ is a rooted ordered tree with each node ν associated to a symbol $s(\nu) \in \mathcal{S}$. The root of $\mathcal{T}(A)$ is a node ρ with $s(\rho) = A$. If $A \in \Sigma$, then ρ has no children. If $A \in \mathcal{N}$ and $\mathsf{rhs}(A) = A_1 \cdots A_a$, then ρ has a children, and the subtree rooted at the *i*th child is (a copy of) $\mathcal{T}(A_i)$. The height height(A) of a symbol $A \in \mathcal{S}$ is defined as the height of its parse tree $\mathcal{T}(A)$. In other words, $\mathsf{height}(A) = 0$ if $A \in \Sigma$ and $\mathsf{height}(A) = 1 + \max_{i=1}^a \mathsf{height}(A_i)$ if $\mathsf{rhs}(A) = A_1 \cdots A_a$. The parse tree $\mathcal{T}_{\mathcal{G}}$ of an SLG \mathcal{G} is defined as the parse tree $\mathcal{T}(S)$ of the starting symbol S, and the height of \mathcal{G} is defined as the height of S.

Each node ν of $\mathcal{T}(A)$ is associated with a fragment $\exp(\nu)$ of $\exp(A)$ matching $\exp(\mathfrak{s}(\nu))$. For the root ρ , we define $\exp(\rho) = \exp(A)[1..|\exp(A)|]$ to be the whole $\exp(A)$. Moreover, if $\exp(\nu) = \exp(A)[\ell..r)$, $\operatorname{rhs}(\mathfrak{s}(\nu)) = A_1 \cdots A_a$, and ν_1, \ldots, ν_a are the children of ν , then $\exp(\nu_i) = \exp(A)[r_{i-1} \ldots r_i)$, where $r_i = \sum_{j=1}^i |\exp(A_j)|$ for $0 \le i \le a$. This way, the fragments $\exp(\nu_i)$ form a partition of $\exp(\nu)$, and $\exp(\nu)$ matches $\exp(\mathfrak{s}(\nu))$ (as claimed).

Without loss of generality, we assume that each symbol $A \in \mathcal{S}$ appears as $s(\nu)$ for a node ν of $\mathcal{T}_{\mathcal{G}}$; the remaining symbols can be removed from \mathcal{G} without affecting the string generated by \mathcal{G} .

Straight-Line Programs We say that a straight-line grammar \mathcal{G} is in Chomsky normal form (CNF) if $|\mathsf{rhs}(A)| = 2$ holds for each $A \in \mathcal{N}$. Such a grammar is also called a straight-line program (SLP). An SLP \mathcal{G} of size g (with g symbols) representing a text T of length n can be stored $\mathcal{O}(g)$ space ($\mathcal{O}(g \log n)$ bits) with each non-terminal $A \in \mathcal{N}$ storing $\mathsf{rhs}(A)$ and $|\exp(A)|$. This representation allows for efficiently traversing the parse tree $\mathcal{T}_{\mathcal{G}}$: given a node ν represented as a pair $(s(\nu), \exp(\nu))$, it is possible to retrieve in constant time an analogous representation of a child ν_i of ν given its index $i \in \{1, 2\}$ (among the children of ν) or an arbitrary position T[j] contained in $\exp(\nu_i)$.

AVL Grammars Rytter [Ryt03] and Charikar et al. [CLL⁺05] provided efficient algorithms that convert the LZ77 parsing of a string into an SLP generating it. The original constructions work only for the non-self-referential version of LZ77, but this restriction has been lifted in subsequent works.

Theorem 5.1 ([Gaw11, KK20b, Theorem 5.1]). Given an LZ77-like parsing of a string T[1..n] into f phrases, an $SLP \mathcal{G}$ of height $\mathcal{O}(\log n)$ and size $\mathcal{O}(f \log n)$ generating T can be constructed in $\mathcal{O}(f \log n)$ time.

The construction in [Ryt03, KK20b] is based on the notion of AVL grammars, which are SLPs satisfying the following extra condition: if $\mathsf{rhs}(A) = BC$ for $A \in \mathcal{N}$, then $|\mathsf{height}(B) - \mathsf{height}(C)| \le 1$. This guarantees [Ryt03, Lemma 1] that $\mathsf{height}(A) = \mathcal{O}(\log|\exp(A)|)$ holds for every $A \in \mathcal{S}$. The algorithm of [Ryt03, KK20b] builds \mathcal{G} incrementally: each step involves adding a symbol A with a desired expansion $\exp(A)$, as well as a bounded number of auxiliary symbols. In the last step, the starting symbol S with $\exp(S) = T$ is added. Each step is of one of three kinds:

- (a) A new terminal symbol can be added to \mathcal{G} in $\mathcal{O}(1)$ time (along with no auxiliary symbols).
- (b) Given two symbols $B, C \in \mathcal{S}$, a new symbol A with $\exp(A) = \exp(B) \exp(C)$ can be added to \mathcal{G} in $\mathcal{O}(1 + \log|\exp(A)|)$ time along with $\mathcal{O}(\log|\exp(A)|)$ auxiliary symbols [Ryt03, Lemma 2].

(c) Given a symbol $A \in \mathcal{S}$ and two positions $1 \leq i \leq j \leq |\exp(A)|$, a new symbol B with $\exp(B) = \exp(A)[i ... j]$ can be added to \mathcal{G} in $\mathcal{O}(1 + \log|\exp(A)|)$ time along with $\mathcal{O}(\log|\exp(A)|)$ auxiliary symbols [Ryt03, Lemma 3 and Theorem 2].

Run-Length Straight-Line Programs A run-length straight-line program (RLSLP) is a straight-line grammar \mathcal{G} whose non-terminals can be classified into pairs with $\mathsf{rhs}(A) = BC$ for symbols $B, C \in \mathcal{S}$ such that $B \neq C$, and powers with $\mathsf{rhs}(A) = B^k$ for a symbol $B \in \mathcal{S}$ and an integer $k \geq 2$. Analogously to an SLP, an RLSLP of size g (with g symbols) representing a text T of length n can be stored in $\mathcal{O}(g)$ space $(\mathcal{O}(g \log n))$ bits) allowing efficient traversal of the parse tree \mathcal{T}_G .

5.2 Run-Length Grammar Construction via Restricted Recompression

Both recompression and restricted recompression, given a string $T \in \Sigma^+$, construct a sequence of strings $(T_k)_{k=0}^{\infty}$ over the alphabet \mathcal{A} defined as the least fixed point of the following equation:

$$\mathcal{A} = \Sigma \cup (\mathcal{A} \times \mathcal{A}) \cup (\mathcal{A} \times \mathbb{Z}_{\geq 2}).$$

Symbols in $\mathcal{A} \setminus \Sigma$ are non-terminals with $\mathsf{rhs}((A_1, A_2)) = A_1 A_2$ for $(A_1, A_2) \in \mathcal{A} \times \mathcal{A}$ and $\mathsf{rhs}((A_1, m)) = A_1^m$ for $(A_1, m) \in \mathcal{A} \times \mathbb{Z}_{\geq 2}$. With any symbol in \mathcal{A} designated as the start symbol, this yields a run-length straight-line program (RLSLP). Intuitively, \mathcal{A} forms a $\mathit{universal}$ RLSLP: for every RLSLP with symbols \mathcal{S} and terminals $\Sigma \subseteq \mathcal{S}$, there is a unique homomorphism $f: \mathcal{S} \to \mathcal{A}$ such that f(A) = A if $A \in \Sigma$ and $\mathsf{rhs}(f(A)) = f(A_1) \cdots f(A_a)$ if $\mathsf{rhs}_{\mathcal{G}}(A) = A_1 \cdots A_a$. As a result, \mathcal{A} provides a convenient formalism to argue about procedures generating RLSLPs.

The main property of strings $(T_k)_{k=0}^{\infty}$ generated using (restricted) recompression is that $\exp(T_k) = T$ holds for all $k \in \mathbb{Z}_{\geq 0}$. The subsequent strings T_k , starting from $T_0 = T$, are obtained by alternate applications of the following two functions which decompose a string of symbols into blocks and then collapse blocks into appropriate symbols. In Definition 5.2, all blocks of length at least 2 are maximal blocks of the same symbol in \mathcal{B} , and they are collapsed to symbols in $\mathcal{B} \times \mathbb{Z}_{\geq 2}$. In Definition 5.4, all blocks consisting of a symbol in \mathcal{L} followed by a symbol in \mathcal{R} are collapsed to a symbol in $\mathcal{L} \times \mathcal{R}$. We provide efficient algorithms that implement both transformations, with the input and the output strings represented using their LZ77 parsings. Previous work [KRRW23, KNP23] relied on straightforward linear-time implementations in the uncompressed settings.

Definition 5.2 (Restricted run-length encoding [KRRW23, KNP23]). Given $T \in \mathcal{A}^+$ and $\mathcal{B} \subseteq \mathcal{A}$, we define $\mathsf{rle}_{\mathcal{B}}(T) \in \mathcal{A}^+$ to be the string obtained as follows by decomposing T into blocks and collapsing these blocks:

- 1. For $i \in [1..|T|)$, place a block boundary between T[i] and T[i+1] unless $T[i] = T[i+1] \in \mathcal{B}$.
- 2. Replace each block $T[i...i+m) = A^m$ of length $m \ge 2$ with a symbol $(A, m) \in \mathcal{A}$.

Lemma 5.3. Given the LZ77 parsing of a string $T \in \mathcal{A}^+$ and a set $\mathcal{B} \subseteq \mathcal{A}$, the LZ77 parsing of $\hat{T} := \mathsf{rle}_{\mathcal{B}}(T)$ can be constructed in $\mathcal{O}(|\mathcal{B}| + z(T)\log^5 n)$ time.

Proof. First, we use Theorem 5.1 to build an SLP \mathcal{G} of size $|\mathcal{G}| = \mathcal{O}(f \log n)$ generating T. Next, we construct a new grammar $\hat{\mathcal{G}}$ (over the same alphabet) of size $\mathcal{O}(|\mathcal{G}|)$ generating \hat{T} . For each symbol A of \mathcal{G} , we construct a non-terminal \hat{A} in $\hat{\mathcal{G}}$, as well as strings LR(A), RR(A) of length 0 or 1 such that:

- $\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(A)) = \mathsf{LR}(A) \cdot \exp_{\hat{\mathcal{G}}}(\hat{A}) \cdot \mathsf{RR}(A);$
- $LR(A) \neq \varepsilon$ if and only if $LS(A) \in \mathcal{B}$;
- $\mathsf{RR}(A) \neq \varepsilon$ if and only if $\mathsf{RS}(A) \in \mathcal{B}$ and $|\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{C}}(A))| > 1$.

It is easy to verify that the following construction satisfies these invariants. For every terminal A of \mathcal{G} , we consider two cases:

- If $A \in \mathcal{B}$, then $\mathsf{LR}(A) = A$, $\mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{A}) = \varepsilon$, and $\mathsf{RR}(A) = \varepsilon$.
- Otherwise, $LR(A) = \varepsilon$, $rhs_{\hat{G}}(\hat{A}) = A$, and $RR(A) = \varepsilon$.

For every non-terminal A with $\mathsf{rhs}_{\mathcal{G}}(A) = BC$, we consider multiple possibilities:

- If $\mathsf{RS}(B) \neq \mathsf{LS}(C)$ or $\mathsf{RS}(B) = \mathsf{LS}(C) \notin \mathcal{B}$, then $\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(A)) = \mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(B)) \cdot \mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(C))$. Thus, we proceed as follows:
 - $-\text{ If } |\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(C))| > 1, \text{ then } \mathsf{LR}(A) = \mathsf{LR}(B), \, \mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{A}) = \hat{B} \cdot \mathsf{RR}(B) \cdot \mathsf{LR}(C) \cdot \hat{C}, \text{ and } \mathsf{RR}(A) = \mathsf{RR}(C). \\ \text{ If } |\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(C))| = 1, \text{ then } \mathsf{LR}(A) = \mathsf{LR}(B), \, \mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{A}) = \hat{B} \cdot \mathsf{RR}(B), \text{ and } \mathsf{RR}(A) = \mathsf{LR}(C).$
- In the remaining cases, $\mathsf{RS}(B) = \mathsf{LS}(C) = X$ for some $X \in \mathcal{B}$. Thus, $\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(A))$ is obtained from $\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(B)) \cdot \mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(C))$ by merging the last run in $\exp_{\mathcal{G}}(B)$ and the first run in $\exp_{\mathcal{G}}(C)$, both of which are powers of X. To formalize this operation, we identify X with (X,1) and define (X,i) + (X,j) := (X,i+j).
 - If $|\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(B))| > 1$ and $|\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(C))| > 1$, then $\mathsf{LR}(A) = \mathsf{LR}(B)$, $\mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{A}) = \hat{B} \cdot (\mathsf{RR}(B) + \mathsf{LR}(C)) \cdot \hat{C}$, and $\mathsf{RR}(A) = \mathsf{RR}(C)$.
 - If $|\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(B))| > 1$ and $|\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(C))| = 1$, then $\mathsf{LR}(A) = \mathsf{LR}(B)$, $\mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{A}) = \hat{B}$, and $\mathsf{RR}(A) = \mathsf{RR}(B) + \mathsf{LR}(C)$.
 - If $|\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(B))| = 1$ and $|\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(C))| > 1$, then $\mathsf{LR}(A) = \mathsf{LR}(B) + \mathsf{LR}(C)$, $\mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{A}) = \hat{C}$, and $\mathsf{RR}(A) = \mathsf{RR}(C)$.
 - If $|\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(B))| = 1$ and $|\mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(C))| = 1$, then $\mathsf{LR}(A) = \mathsf{LR}(B) + \mathsf{LR}(C)$, $\mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{A}) = \varepsilon$, and $\mathsf{RR}(A) = \varepsilon$.

Additionally, we add to $\hat{\mathcal{G}}$ a new starting symbol \hat{S}' such that $\mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{S}') = \mathsf{LR}(S) \cdot \hat{S} \cdot \mathsf{RR}(S)$, where S is the starting symbol of \mathcal{G} . Thus, the grammar $\hat{\mathcal{G}}$ is of size $|\hat{\mathcal{G}}| = \mathcal{O}(|\mathcal{G}|)$ and generates $\exp_{\hat{\mathcal{G}}}(\hat{S}') = \mathsf{rle}_{\mathcal{B}}(\exp_{\mathcal{G}}(S)) = \mathsf{rle}_{\mathcal{B}}(T) = \hat{T}$. We construct an LZ77-like parsing of \hat{T} of size $\hat{\mathcal{G}}$ by traversing the parse tree of $\hat{\mathcal{G}}$ and creating a new previous factor whenever the symbol $\mathsf{s}(\nu)$ has already been encountered, This parsing can be converted into the LZ77 parsing of \hat{T} in $\mathcal{O}(|\hat{\mathcal{G}}|\log^4 n) = \mathcal{O}(z(T)\log^5 n)$ time using Proposition 2.3.

Definition 5.4 (Restricted pair compression [KRRW23, KNP23]). Given $T \in \mathcal{A}^+$ and disjoint sets $\mathcal{L}, \mathcal{R} \subseteq \mathcal{A}$, we define $\mathsf{pc}_{\mathcal{L},\mathcal{R}}(T) \in \mathcal{A}^+$ to be the string obtained as follows by decomposing T into blocks and collapsing these blocks:

- 1. For $i \in [1..|T|)$, place a block boundary between T[i] and T[i+1] unless $T[i] \in \mathcal{L}$ and $T[i+1] \in \mathcal{R}$.
- 2. Replace each block T[i...i+1] of length 2 with a symbol $(T[i], T[i+1]) \in \mathcal{A}$.

Lemma 5.5. Given the LZ77 parsing of a string $T \in \mathcal{A}^n$ and disjoint sets $\mathcal{L}, \mathcal{R} \subseteq \mathcal{A}$, the LZ77 parsing of $\hat{T} := \mathsf{pc}_{\mathcal{L},\mathcal{R}}(T)$ can be constructed in $\mathcal{O}(|\mathcal{L}| + |\mathcal{R}| + z(T)\log^5 n)$ time.

Proof. First, we use Theorem 5.1 to build an SLP \mathcal{G} of size $\mathcal{O}(f \log n)$ generating T. We construct a new grammar $\hat{\mathcal{G}}$ (over the same alphabet) of size $\mathcal{O}(|\mathcal{G}|)$ generating \hat{T} . For each symbol X of \mathcal{G} , we construct a non-terminal \hat{A} in $\hat{\mathcal{G}}$, as well as strings LB(A), RB(A) of length 0 or 1 such that:

- $\operatorname{pc}_{\mathcal{L},\mathcal{R}}(\exp_{\mathcal{G}}(A)) = \operatorname{LB}(A) \cdot \exp_{\hat{\mathcal{G}}}(\hat{A}) \cdot \operatorname{RB}(A);$
- $\mathsf{LB}(A) \neq \varepsilon$ if and only if $\mathsf{LS}(A) \in \mathcal{R}$;
- $\mathsf{RB}(A) \neq \varepsilon$ if and only if $\mathsf{RS}(A) \in \mathcal{L}$.

It is easy to verify that the following construction satisfies these invariants. For every terminal A of \mathcal{G} , we consider two cases:

- If $A \in \mathcal{L}$, then $\mathsf{LB}(A) = \varepsilon$, $\mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{A}) = \varepsilon$, and $\mathsf{RB}(A) = A$.
- If $A \in \mathcal{R}$, then $\mathsf{LB}(A) = A$, $\mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{A}) = \varepsilon$, and $\mathsf{RB}(A) = \varepsilon$.
- Otherwise, $\mathsf{LB}(A) = \varepsilon$, $\mathsf{rhs}_{\hat{G}}(\hat{A}) = A$, and $\mathsf{RB}(A) = \varepsilon$.

For every non-terminal A with $\exp_{\mathcal{G}}(A) = BC$, we consider two possibilities:

- If $\mathsf{RB}(B) = \varepsilon$ or $\mathsf{LB}(C) = \varepsilon$, then $\mathsf{pc}_{\mathcal{L},\mathcal{R}}(\exp_{\mathcal{G}}(A)) = \mathsf{pc}_{\mathcal{L},\mathcal{R}}(\exp_{\mathcal{G}}(B)) \cdot \mathsf{pc}_{\mathcal{L},\mathcal{R}}(\exp_{\mathcal{G}}(C))$. Thus, we set $\mathsf{LB}(A) = \mathsf{LB}(B)$, $\mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{A}) = \hat{B} \cdot \mathsf{RB}(B) \cdot \mathsf{LB}(C) \cdot \hat{C}$, and $\mathsf{RB}(A) = \mathsf{RB}(C)$.
- Otherwise, $\operatorname{pc}_{\mathcal{L},\mathcal{R}}(\exp_{\mathcal{G}}(A))$ is obtained from $\operatorname{pc}_{\mathcal{L},\mathcal{R}}(\exp_{\mathcal{G}}(B)) \cdot \operatorname{pc}_{\mathcal{L},\mathcal{R}}(\exp_{\mathcal{G}}(C))$ by merging the last symbol of B and the first symbol of C into a single block $(\operatorname{RB}(B),\operatorname{LB}(C))$. Consequently, we set $\operatorname{LB}(A) = \operatorname{LB}(B)$, $\operatorname{rhs}_{\hat{G}}(\hat{A}) = \hat{B} \cdot (\operatorname{RB}(B),\operatorname{LB}(C)) \cdot \hat{C}$, and $\operatorname{RB}(A) = \operatorname{RB}(C)$.

Additionally, we add to $\hat{\mathcal{G}}$ a new starting symbol \hat{S}' such that $\hat{S}' \to \mathsf{LB}(S) \cdot \hat{S} \cdot \mathsf{RB}(S)$, where S is the starting symbol of \mathcal{G} . Thus, the grammar $\hat{\mathcal{G}}$ is of size $|\hat{\mathcal{G}}| = \mathcal{O}(|\mathcal{G}|)$ and generates $\exp_{\hat{\mathcal{G}}}(\hat{S}') = \mathsf{pc}_{\mathcal{L},\mathcal{R}}(\exp_{\mathcal{G}}(S)) = \mathsf{pc}_{\mathcal{L},\mathcal{R}}(\exp_{\mathcal{G}}(S))$ $\operatorname{pc}_{\mathcal{L},\mathcal{R}}(T) = \hat{T}$. We construct an LZ77-like parsing of \hat{T} of size $\hat{\mathcal{G}}$ by traversing the parse tree of $\hat{\mathcal{G}}$ and creating a new previous factor whenever the symbol $s(\nu)$ has already been encountered, This parsing can be converted into the LZ77 parsing of \hat{T} in $\mathcal{O}(|\hat{\mathcal{G}}|\log^4 n) = \mathcal{O}(z(T)\log^5 n)$ time using Proposition 2.3.

We are now ready to formally define the sequence $(T_k)_{k=0}^{\infty}$ constructed through restricted recompression.

Construction 5.6 (Restricted recompression [KRRW23, KNP23]). Given a string $T \in \Sigma^+$, the strings T_k for $k \in \mathbb{Z}_{\geq 0}$ are constructed as follows, based on $\ell_k := (\frac{8}{7})^{\lceil \frac{k}{2} \rceil - 1}$ and $\mathcal{A}_k := \{A \in \mathcal{A} : |\exp(A)| \leq \ell_k\}$:

- If k = 0, then $T_k = T$.
- If k > 0 is odd, then $T_k = \mathsf{rle}_{\mathcal{A}_k}(T_{k-1})$. If k > 0 is even, then $T_k = \mathsf{pc}_{\mathcal{L}_k, \mathcal{R}_k}(T_{k-1})$, where $\mathcal{L}_k, \mathcal{R}_k \subseteq \mathcal{A}_k$ are disjoint.

It is easy to see that $\exp(T_k) = T$ indeed holds for all $k \in \mathbb{Z}_{>0}$. The main challenge is to appropriately select the subsets $\mathcal{L}_k, \mathcal{R}_k$. The first (easier) goal is to make sure that $|T_k| = 1$ holds for some $k \in \mathbb{Z}_{\geq 0}$ so that an RLSLP generating T can be obtained by setting $T_k[1]$ as the starting symbol of the RLSLP derived from by A. While this RLSLP contains infinitely many symbols, we can remove symbols that do not occur in any string T_k . Formally, for each $k \in \mathbb{Z}_{>0}$, we define the family $S_k := \{T_k[j] : j \in [1..|T_k|]\} \subseteq A$ of symbols occurring in T_k . Then, the actual symbols present in the generated RLSLP can be expressed as $\mathcal{S} := \bigcup_{k=0}^{\infty} \mathcal{S}_k$.

The main goal behind our selection of the subsets $\mathcal{L}_k, \mathcal{R}_k$ is to make sure that $|\mathcal{S}| = \mathcal{O}(\delta(T) \cdot \frac{\tilde{n} \log \sigma}{\delta(T) \log n})$. As shown in [KNP23], this holds in expectation if \mathcal{L}_k and \mathcal{R}_k form a random partition of \mathcal{A}_k . In this work, we develop an alternative deterministic construction parameterized by a cover hierarchy C of the string T (cf. Definition 4.3)

Construction 5.7. Fix a text $T \in \Sigma^n$ and its cover hierarchy C. For every $k \in \mathbb{Z}_{\geq 0}$, denote $\alpha_k = \lfloor 16\ell_k \rfloor$ and $m_k = 2\alpha_k + \lfloor \ell_{k+1} \rfloor$, For integers $k, h \in \mathbb{Z}_{>0}$, let us define

$$J_{k,h} := \{ j \in [1 ... | T_{k-1}|) : | \exp(T_{k-1}[1 ... j])| \in \mathsf{C}_{m_h} \},$$

$$w_{k,h}(A,B) := | \{ j \in J_{k,h} : T_{k-1}[j ... j+1] = AB \} |,$$

$$w_k(A,B) := \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} w_{k,h}(A,B).$$

The subsets \mathcal{L}_k , $\mathcal{R}_k \subseteq \mathcal{A}_k$ are chosen so that $\sum_{(A,B)\in\mathcal{L}_k\times\mathcal{R}_k} w_k(A,B) \ge \frac{1}{4} \sum_{(A,B)\in\mathcal{A}_k\times\mathcal{A}_k} w_k(A,B)$.

5.2.1 Analysis of the Grammar Size

Our argument relies on several properties of restricted recompression proved in [KRRW23, KNP23].

Fact 5.8 ([KNP23, Fact V.7]). For every $k \in \mathbb{Z}_{>0}$, if $\exp(x) = \exp(x')$ holds for two fragments of T_k , then x = x'.

Corollary 5.9 ([KNP23, Corollary V.8]). For every odd $k \in \mathbb{Z}_{>0}$, there is no $j \in [1..|T_k|)$ such that $T_k[j] = T_k[j+1] \in \mathcal{A}_{k+1}$.

Recall that $\exp(T_k) = T$ for every $k \in \mathbb{Z}_{\geq 0}$. Hence, for every $j \in [1..|T_k|]$, we can associate $T_k[j]$ with a fragment $T(|\exp(T_k[1..j))|..|\exp(T_k[1..j))|] = \exp(T_k[j])$; these fragments are called *phrases* (of T) induced by T_k . We also define a set B_k of phrase boundaries induced by T_k :

$$\mathsf{B}_k = \{ |\exp(T_k[1\mathinner{.\,.} j])| : j \in [1\mathinner{.\,.} |T_k|) \}.$$

Fact 5.10 (see [KRRW23]). For every $k \in \mathbb{Z}_{>0}$ and every symbol $A \in \mathcal{S}_k$, the string $\exp(A)$ has length at $most \ 2\ell_k \ or \ period \ at \ most \ \ell_k.$

Proof. We proceed by induction on k. Let $A \in \mathcal{S}_k$. If k = 0, then $|\exp(A)| = 1 \le 2 \cdot \frac{7}{8} = 2\ell_0$. Thus, we may assume k > 0. If $A \in \mathcal{S}_{k-1}$, then the inductive assumption shows that $\exp(A)$ is of length at most $2\ell_{k-1} \le 2\ell_k$ or period most $\ell_{k-1} \le \ell_k$. Otherwise, we have two possibilities. If k is odd, then $A = (B, m) \in \mathcal{B}_k \times \mathbb{Z}_{\ge 2}$, and thus the period $\exp(A)$ is at most $|\exp(B)| \le \ell_k$. If k is even, on the other hand, then $A = (B, C) \in \mathcal{B}_k^2$, so $|\exp(A)| = |\exp(B)| + |\exp(C)| \le 2\ell_k$.

The following lemma captures the "local consistency" property of our construction: the presence of phrase boundaries is determined by a small context.

Lemma 5.11 ([KNP23, Lemma V.9]). Let $\alpha \in \mathbb{Z}_{\geq 1}$ and let $i, i' \in [\alpha ... n - \alpha]$ be such that $T(i - \alpha ... i + \alpha] = T(i' - \alpha ... i' + \alpha]$. For every $k \in \mathbb{Z}_{>0}$, if $\alpha \geq \lfloor 16\ell_k \rfloor$, then $i \in \mathsf{B}_k \iff i' \in \mathsf{B}_k$.

As a first step towards proving $|S| = \mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$, we apply Lemma 5.11 to bound $|S_k \setminus S_{k+1}|$ for a given $k \in \mathbb{Z}_{>0}$.

Lemma 5.12. For every $k \in \mathbb{Z}_{\geq 0}$, we have $|S_k \setminus S_{k+1}| \leq 2|B_k \cap C_{m_k}|$.

Proof. First, suppose that $|\mathsf{B}_k \cap \mathsf{C}_{m_k}| = 0$. Consider $p = \min \mathsf{B}_k$. Since $p \notin \mathsf{C}_{m_k}$, we must have $p > m_k$ and $p - \alpha_k + 1 \neq \mathsf{Occ}_{m_k}(p - \alpha_k + 1, T)$. Consequently, there is a position $p' \in [\alpha_k \dots p)$ such that $T(p - \alpha_k \dots p - \alpha_k + m_k] = T(p' - \alpha_k \dots p' - \alpha_k + m_k]$. In particular, $T(p - \alpha_k \dots p + \alpha_k] = T(p' - \alpha_k \dots p' + \alpha_k]$, so Lemma 5.11 yields $p' \in \mathsf{B}_k$, contradicting the choice of p. Thus, $\mathsf{B}_k = \emptyset$, which means that $\mathcal{S}_{k+1} = \mathcal{S}_k = \{T_k[1]\}$.

Next, we prove that $|\mathcal{S}_k \setminus \mathcal{S}_{k+1}| \leq 1 + |\mathsf{B}_k \cap \mathsf{C}_{m_k}|$. Let $T_k[j]$ be the leftmost occurrence in T_k of $A \in \mathcal{S}_k \setminus \mathcal{S}_{k+1}$. Moreover, let $p = |\exp(T_k[1..j])|$ and $q = |\exp(T_k[1..j])|$ so that $T(p..q] = \exp(A)$ is the phrase induced by $T_k[j]$. By Construction 5.6, we have $A \in \mathcal{A}_{k+1}$, and therefore $q - p \leq \ell_{k+1}$.

We shall prove that j=1 or $p \in B_k \cap C_{m_k}$. This will complete the proof of $|S_k \setminus S_{k+1}| \le 1 + |B_k \cap C_{m_k}|$ because distinct symbols A yield distinct positions j and p. For a proof by contradiction, suppose that $j \in (1..|T_k|]$ yet $p \notin B_k \cap C_{m_k}$. Since $p \in B_k$ holds due to j > 1, we derive $p \notin C_{m_k}$. Hence, $p > m_k$ and $p - \alpha_k + 1 \ne Occ_{m_k}(p - \alpha_k + 1, T)$. Consequently, there is a position $p' \in [\alpha_k ... p)$ such that $T(p - \alpha_k ... p - \alpha_k + m_k] = T(p' - \alpha_k ... p' - \alpha_k + m_k]$. In particular, $T(p - \alpha_k ... p + \alpha_k] = T(p' - \alpha_k ... p' + \alpha_k]$, so Lemma 5.11 yields $p' \in B_k$. Similarly, due to $q - p = |\exp(A)| \le \lfloor \ell_{k+1} \rfloor = m_k - 2\alpha_k$, we have $T(q - \alpha_k ... q + \alpha_k] = T(q' - \alpha_k ... q' + \alpha_k]$ for $q' := p' + |\exp(A)|$, and therefore $q' \in B_k$ holds due to $q \in B_k$. Lemma 5.11 further implies $B_k \cap (p' ... q') = \emptyset = B_k \cap (p ... q)$. Consequently, T(p' ... q') is a phrase induced by T_k , and, since p' < p, it corresponds to $T_k[j']$ for some j' < j. By Fact 5.8, we have $T_k[j'] = T_k[j] = A$, which contradicts the choice of $T_k[j]$ as the leftmost occurrence of A in T_k .

Overall, we have $|\mathcal{S}_k \setminus \mathcal{S}_{k+1}| = 0$ if $|\mathsf{B}_k \cap \mathsf{C}_{m_k}| = 0$ and $|\mathcal{S}_k \setminus \mathcal{S}_{k+1}| \le 1 + |\mathsf{B}_k \cap \mathsf{C}_{m_k}|$ otherwise. Combining these two claims, we derive $|\mathcal{S}_k \setminus \mathcal{S}_{k+1}| \le 2|\mathsf{B}_k \cap \mathsf{C}_{m_k}|$.

Next, we use our choice of \mathcal{L}_k and \mathcal{R}_k to bound the sizes $|\mathsf{B}_h \cap \mathsf{C}_{m_h}|$ in terms of the sizes $|\mathsf{C}_{m_h}|$.

Lemma 5.13. For every $k \in \mathbb{Z}_{\geq 0}$, we have

$$\sum_{h=0}^{k-1} |\mathsf{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} |\mathsf{B}_k \cap \mathsf{C}_{m_h}| \leq \sum_{h=0}^{k-1} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{h+1}} \right) + \sum_{h=k}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{k+1}} \right).$$

Proof. Denote

$$L_k := \sum_{h=0}^{k-1} |\mathsf{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} |\mathsf{B}_k \cap \mathsf{C}_{m_h}|.$$

We proceed by induction on k. For k=0, due to $\ell_1=1$, we have

$$L_0 = \sum_{h=0}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor} |\mathsf{B}_0 \cap \mathsf{C}_{m_h}|$$

$$\begin{split} &\leq \sum_{h=0}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor} |\mathsf{C}_{m_h}| \\ &< \sum_{h=0}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor} \left(\tfrac{2|\mathsf{C}_{m_h}|}{m_h} + \tfrac{4|\mathsf{C}_{m_h}|}{\ell_1} \right). \end{split}$$

If k is odd, then $\mathsf{B}_k \subseteq \mathsf{B}_{k-1}, \lfloor \frac{k}{2} \rfloor = \lfloor \frac{k-1}{2} \rfloor$, and $\ell_{k+1} = \ell_k$. Therefore,

$$\begin{split} L_k &= \sum_{h=0}^{k-1} |\mathsf{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} |\mathsf{B}_k \cap \mathsf{C}_{m_h}| \\ &\leq \sum_{h=0}^{k-1} |\mathsf{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k-1}{2} \rfloor} |\mathsf{B}_{k-1} \cap \mathsf{C}_{m_h}| \\ &= \sum_{h=0}^{k-2} |\mathsf{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k-1}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k-1}{2} \rfloor} |\mathsf{B}_{k-1} \cap \mathsf{C}_{m_h}| \\ &\leq \sum_{h=0}^{k-2} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{h+1}} \right) + \sum_{h=k-1}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k-1}{2} \rfloor} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_k} \right) \\ &= \sum_{h=0}^{k-1} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{h+1}} \right) + \sum_{h=k}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{k+1}} \right). \end{split}$$

It remains to consider the case when k is even.

Claim 5.14. For every $h \in \mathbb{Z}_{\geq 0}$, the set $J'_{k,h} := \{j \in J_{k,h} : T_{k-1}[j] \notin \mathcal{A}_k \text{ or } T_{k-1}[j+1] \notin \mathcal{A}_k \}$ satisfies $|J'_{k,h}| < \frac{2}{m_h} |\mathsf{C}_{m_h}| + \frac{2}{\ell_k} |\mathsf{C}_{m_h}|$.

Proof. Let us first consider the case when $m_h \ge n$ so that $\mathsf{C}_{m_h} = [1 \dots n]$. If $j \in J'_{k,h}$, then $T_{k-1}[j] \notin \mathcal{A}_k$ or $T_{k-1}[j+1] \notin \mathcal{A}_k$. The level-(k-1) phrase corresponding to this pausing symbol is of length greater than ℓ_k , and there are fewer than $\frac{n}{\ell}$ such phrases. Thus, $|J'_{k,h}| < \frac{2n}{\ell} < \frac{2}{\ell} - |\mathsf{C}_{m_k}| + \frac{2}{\ell} |\mathsf{C}_{m_k}|$.

and there are fewer than $\frac{n}{\ell_k}$ such phrases. Thus, $|J'_{k,h}| < \frac{2n}{\ell_k} < \frac{1}{2m} | \mathsf{C}_{m_h} | \mathsf{C}_{m_h} | + \frac{2}{\ell_k} | \mathsf{C}_{m_h} |$. Next, consider $m_h < n$. Denote $\bar{\mathsf{B}}_{k-1} = \mathsf{B}_{k-1} \cup \{0,n\}$. Let us fix an interval $I \subseteq \mathsf{C}_{m_h}$ and define $J = \{j \in [1..|T_{k-1}|) : |\exp(T_{k-1}[1..j-1])| \in I \text{ and } |\exp(T_{k-1}[1..j+1])| \in I\}$. Observe that $|J| = \max(0,|\bar{\mathsf{B}}_{k-1}\cap I|-2)$: every $j\in J$ corresponds to a position $|\exp(T_{k-1}[1..j])| \in \bar{\mathsf{B}}_{k-1}\cap I$ which is neither the leftmost nor the rightmost one in $\bar{\mathsf{B}}_{k-1}\cap I$. If $j\in J'_{k,h}$, then $T_{k-1}[j]\notin \mathcal{A}_k$ or $T_{k-1}[j+1]\notin \mathcal{A}_k$. The level-(k-1) phrase T(p..q] corresponding to this pausing symbol satisfies $[p..q]\cap I$ and, by definition of \mathcal{A}_k , it is of length $q-p>\ell_k$. There are fewer than $\frac{|I|-1}{\ell_k}$ phrases T(p..q] satisfying both conditions, and thus $|J\cap J'_{k,h}|<\frac{2|I|-2}{\ell_k}$. Taking into account the leftmost and the rightmost element of $\bar{\mathsf{B}}_{k-1}\cap I$, we conclude that $|\{j\in J'_{k,h}| < \frac{2|I|-2}{\ell_k}\}$. Taking into account the leftmost and the rightmost element of $\bar{\mathsf{B}}_{k-1}\cap I$, we conclude that $|\{j\in J'_{k,h}| < \frac{2|C_{m_h}|}{\ell_k}\}$. Since $|\mathcal{I}(C_{m_h})| \leq \frac{1}{m_h}|C_{m_h}|$, we conclude that $|J'_{k,h}| < \frac{2|C_{m_h}|}{m_h} + \frac{2|C_{m_h}|}{\ell_k}$.

Therefore,

$$\begin{split} L_k &= \sum_{h=0}^{k-1} |\mathsf{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} |\mathsf{B}_k \cap \mathsf{C}_{m_h}| \\ &= \sum_{h=0}^{k-1} |\mathsf{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} (|\mathsf{B}_{k-1} \cap \mathsf{C}_{m_h}| - |\{j \in J_{k,h} : T_{k-1}[j \mathinner{\ldotp\ldotp\ldotp} j+1] \in \mathcal{L}_k \mathcal{R}_k\}|) \\ &= \sum_{h=0}^{k-1} |\mathsf{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\tfrac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \left(|\mathsf{B}_{k-1} \cap \mathsf{C}_{m_h}| - \sum_{(A,B) \in \mathcal{L}_k \times \mathcal{R}_k} w_{k,h}(A,B) \right) \end{split}$$

$$\begin{split} &= \sum_{h=0}^{k-1} |\mathcal{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} |\mathcal{B}_{k-1} \cap \mathsf{C}_{m_h}| - \sum_{(A,B) \in \mathcal{L}_k \times \mathcal{R}_k} w_k(A,B) \\ &\leq \sum_{h=0}^{k-1} |\mathcal{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} |\mathcal{B}_{k-1} \cap \mathsf{C}_{m_h}| - \frac{1}{4} \sum_{(A,B) \in \mathcal{A}_k \times \mathcal{A}_k} w_k(A,B) \\ &= \sum_{h=0}^{k-1} |\mathcal{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \left(|\mathcal{B}_{k-1} \cap \mathsf{C}_{m_h}| - \frac{1}{4} \sum_{(A,B) \in \mathcal{A}_k \times \mathcal{A}_k} w_{k,h}(A,B) \right) \\ &= \sum_{h=0}^{k-1} |\mathcal{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} (|\mathcal{B}_{k-1} \cap \mathsf{C}_{m_h}| - \frac{1}{4} |\{j \in J_{k,h} : T_{k-1}[j ... j+1] \in \mathcal{A}_k \mathcal{A}_k\}|) \\ &= \sum_{h=0}^{k-1} |\mathcal{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} (\frac{3}{4} |\mathcal{B}_{k-1} \cap \mathsf{C}_{m_h}| + \frac{1}{4} |J'_{k,h}|) \\ &= \sum_{h=0}^{k-2} |\mathcal{B}_h \cap \mathsf{C}_{m_h}| + \sum_{h=k-1}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k-1}{2} \rfloor} |\mathcal{B}_{k-1} \cap \mathsf{C}_{m_h}| + \frac{1}{4} \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} |J'_{k,h}| \\ &\leq \sum_{h=0}^{k-2} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{h+1}} \right) + \sum_{h=k-1}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_k} \right) + \frac{1}{4} \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_k} \right) \\ &= \sum_{h=0}^{k-1} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{h+1}} \right) + \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{7|\mathsf{C}_{m_h}|}{\ell_{k+1}} \right) \\ &= \sum_{h=0}^{k-1} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{h+1}} \right) + \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{k+1}} \right) . \\ &= \sum_{h=0}^{k-1} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{h+1}} \right) + \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{k+1}} \right) . \\ &= \sum_{h=0}^{k-1} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}{\ell_{h+1}} \right) + \sum_{h=k}^{\infty} (\frac{3}{4})^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \left(\frac{2|\mathsf{C}_{m_h}|}{m_h} + \frac{4|\mathsf{C}_{m_h}|}$$

This completes the proof.

The following result is used to conclude that $|T_k| = 1$ for sufficiently large $k = \mathcal{O}(\log n)$.

Lemma 5.15. Let $\kappa = 2\lceil \log_{8/7} n \rceil$. For every integer $k \geq \kappa$, we have $|\mathsf{B}_k| < (\frac{3}{4})^{\lfloor \frac{k}{2} \rfloor - \lfloor \frac{\kappa}{2} \rfloor} n$.

Proof. We proceed by induction on n. The base case of $k=\kappa$ is trivial because $\mathsf{B}_\kappa\subseteq[1\mathinner{\ldotp\ldotp} n)$. If $k>\kappa$ is odd, then $|\mathsf{B}_k|\leq |\mathsf{B}_{k-1}|<(\frac34)^{\lfloor\frac{k-1}2\rfloor-\lfloor\frac{\kappa}2\rfloor}n=(\frac34)^{\lfloor\frac{k}2\rfloor-\lfloor\frac{\kappa}2\rfloor}n$. If $k>\kappa$ is even, then we note that $\ell_k\geq n$ and $\mathsf{C}_{m_h}=[1\mathinner{\ldotp\ldotp} n]$ for all $h\geq k$. Consequently, $J_{k,h}=[1\mathinner{\ldotp\ldotp} l]$ and, for every $A,B\in\mathcal{A}$, we have

$$w_k(A, B) = \sum_{k=h}^{\infty} \left(\frac{3}{4}\right)^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} |\{j \in J_{k,h} : T_{k-1}[j \dots j+1] = AB\}|$$

$$= \sum_{k=h}^{\infty} \left(\frac{3}{4}\right)^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} |\{j \in [1 \dots | T_{k-1}|) : T_{k-1}[j \dots j+1] = AB\}|$$

$$= 8|\{j \in [1 \dots | T_{k-1}|) : T_{k-1}[j \dots j+1] = AB\}|$$

Consequently,

$$\begin{aligned} |\mathsf{B}_k| & \leq |\mathsf{B}_{k-1}| - |\{j \in [1\mathinner{\ldotp\ldotp} |T_{k-1}|) : T_{k-1}[j\mathinner{\ldotp\ldotp} j+1] \in \mathcal{L}_k \cdot \mathcal{R}_k\}| \\ & = |\mathsf{B}_{k-1}| - \tfrac{1}{8} \sum_{(A,B) \in \mathcal{L}_k \times \mathcal{R}_k} w_k(A,B) \leq |\mathsf{B}_{k-1}| - \tfrac{1}{32} \sum_{(A,B) \in \mathcal{A}_k \times \mathcal{A}_k} w_k(A,B) = |\mathsf{B}_{k-1}| - \tfrac{1}{4} |\mathsf{B}_{k-1}| \end{aligned}$$

$$= \frac{3}{4} |\mathsf{B}_{k-1}| < \frac{3}{4} \cdot (\frac{3}{4})^{\left\lfloor \frac{k-1}{2} \right\rfloor - \left\lfloor \frac{\kappa}{2} \right\rfloor} n = (\frac{3}{4})^{\left\lfloor \frac{k}{2} \right\rfloor - \left\lfloor \frac{\kappa}{2} \right\rfloor} n. \quad \Box$$

Corollary 5.16. If $k \ge 2\lceil \log_{8/7} n \rceil + 2\lceil \log_{4/3} n \rceil$, then $|T_k| = 1$.

We conclude the analysis of |S| by combining Lemmas 5.12 and 5.13.

Theorem 5.17. We have $|S| = \mathcal{O}(\sum_{k=0}^{\infty} \frac{1}{m_k} \cdot |C_{m_k}|)$.

Proof. By Corollary 5.16, there exists $k = \mathcal{O}(\log n)$ such that $|T_k| = 1$ (and thus $T_h = T_k$ for all $h \ge k$). Applying Lemmas 5.12 and 5.13 for this k, we conclude that

$$\begin{split} |\mathcal{S}| &= |\mathcal{S}_k| + \sum_{h=0}^{k-1} |\mathcal{S}_h \setminus \mathcal{S}_{h+1}| \\ &\leq 1 + 2 \sum_{h=0}^{k-1} |\mathcal{B}_h \cap \mathcal{C}_{m_h}| \\ &\leq 1 + 2 \sum_{h=0}^{k-1} \left(\frac{2|\mathcal{C}_{m_h}|}{m_h} + \frac{4|\mathcal{C}_{m_h}|}{\ell_{h+1}} \right) + 2 \sum_{h=k}^{\infty} \left(\frac{3}{4} \right)^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \cdot \left(\frac{2|\mathcal{C}_{m_h}|}{m_h} + \frac{4|\mathcal{C}_{m_h}|}{\ell_{k+1}} \right) \\ &\leq 1 + 2 \sum_{h=0}^{k-1} \left(\frac{2|\mathcal{C}_{m_h}|}{m_h} + \frac{4|\mathcal{C}_{m_h}|}{\ell_{h+1}} \right) + 2 \sum_{h=k}^{\infty} \left(\frac{7}{8} \right)^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} \cdot \left(\frac{2|\mathcal{C}_{m_h}|}{m_h} + \frac{4|\mathcal{C}_{m_h}|}{\ell_{k+1}} \right) \\ &\leq 1 + 2 \sum_{h=0}^{\infty} \left(\frac{2|\mathcal{C}_{m_h}|}{m_h} + \frac{4|\mathcal{C}_{m_h}|}{\ell_{h+1}} \right) \\ &\leq 1 + 268 \sum_{h=0}^{\infty} \frac{|\mathcal{C}_{m_h}|}{m_h} \\ &= \mathcal{O}\left(\sum_{h=0}^{\infty} \frac{|\mathcal{C}_{m_h}|}{m_h} \right). \end{split}$$

Here, the last inequality follows from the fact that $m_h = \lfloor 16\ell_k \rfloor + \lfloor \ell_{k+1} \rfloor \leq 33\ell_{k+1}$.

Finally, we provide a concrete bound for the cover hierarchy of Construction 4.6.

Lemma 5.18. The cover hierarchy C of Construction 4.6 satisfies $\sum_{k=0}^{\infty} \frac{1}{m_k} |C_{m_k}| = \mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$.

Proof. Define $\mu = \lfloor 2 \log_{8/7} \frac{\log \delta(T)}{264 \log \sigma} \rfloor$ and $\nu = \lceil 2 \log_{8/7} \frac{n}{\delta(T)} \rceil$ so that

$$m_{\mu} \leq 33\ell_{\mu+1} = 33(\frac{8}{7})^{\lfloor \mu/2 \rfloor} \leq 33(\frac{8}{7})^{\log_{8/7}} \frac{\log \delta(T)}{264 \log \sigma} = \frac{\log \delta(T)}{8 \log \sigma},$$

$$\ell_{\nu+2} = (\frac{8}{7})^{\lceil \nu/2 \rceil} \geq (\frac{8}{7})^{\log_{8/7}} \frac{n}{\delta(T)} = \frac{n}{\delta(T)}.$$

For $k \in [0..\mu]$, we observe that $|\mathsf{C}_{m_k}| \leq \mathsf{d}_{8m_k}(T) + 8m_k \leq \sigma^{8m_k} + 8m_k \leq 2\sigma^{8m_k}$. Thus,

$$\sum_{k=0}^{\mu} \frac{|\mathsf{C}_{m_k}|}{m_k} \le \sum_{k=0}^{\mu} |\mathsf{C}_{m_k}| \le \sum_{k=0}^{\mu} 2\sigma^{8m_k} \le 2\sigma^{8m_{\mu}} \le 2\delta(T).$$

For $k \in (\mu ... \nu]$, we observe that $|C_{m_k}| \leq d_{8m_k}(T) + 8m_k \leq 16\delta(T)m_k$. Thus,

$$\sum_{k=\mu+1}^{\nu} \frac{|\mathsf{C}_{m_k}|}{m_k} \leq \sum_{k=\mu+1}^{\nu} 16\delta(T) = 16\delta(T)(\nu - \mu) = \mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n}) = \mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n}).$$

For $k \in (\nu ... \infty)$, we observe that $|C_{m_k}| \leq n$. Thus,

$$\sum_{k=\nu+1}^{\infty} \frac{|\mathsf{C}_{m_k}|}{m_k} \le \sum_{k=\nu+1}^{\infty} \frac{n}{\ell_{k+1}} \le \frac{2n}{\ell_{\nu+2}} \sum_{i=0}^{\infty} (\frac{7}{8})^i \le 8\delta(T).$$

5.3 Efficient Construction

Our next big goal is to develop an efficient implementation of Construction 5.6. We start with an auxiliary tools needed to find the sets \mathcal{L}_k and \mathcal{R}_k according to Construction 5.7. The first of them is a folklore linear-time approximation algorithm for the directed max-cut problem.

Lemma 5.19. There exists a linear-time algorithm that, given a weighted directed graph G = (V, E, w) without self-loops, constructions a partition $V = L \cup R$ such that $w(L, R) \ge \frac{1}{4}w(V, V)$, where, for any $A, B \subseteq V$, we denote $w(A, B) = \sum_{e \in E \cap (A \times B)} w(e)$.

Proof. First, we preprocess G so that each $v \in V$ stores both incoming and outgoing arcs. We maintain a partition $V = L \cup M \cup R$ into three disjoint classes. Initially, M = V and, until $M \neq \emptyset$, we pick an arbitrary vertex $v \in M$ and move v to L or R, depending on whether $2w(v,R) + w(v,M) \ge 2w(L,v) + w(M,v)$ or not. This decision can be implemented in $\mathcal{O}(1 + \deg(v))$ time, which yields a total running time of $\mathcal{O}(|V| + |E|)$.

As for correctness, we shall prove that $\Phi := 4w(L,R) + 2w(L,M) + 2w(M,R) + w(M,M)$ cannot decrease throughout the algorithm. Consider the effect of moving v from M to L on the four terms of Φ (recall that there are no self-loops incident to v):

- w(L,R) increases by w(v,R);
- w(L, M) increases by w(v, M) and decreases by w(L, v);
- w(M,R) decreases by w(v,R);
- w(M, M) decreases by w(v, M) and decreases by w(M, v).

Overall, Φ increases by

$$4w(v,R) + 2(w(v,M) - w(L,v)) - 2w(v,R) - (w(v,M) + w(M,v))$$

$$= 2w(v,R) + w(v,M) - 2w(L,v) - w(M,v),$$

and this quantity is nonnegative when the algorithm decides to move v to L.

Similarly, if v is moved from M to R, then

- w(L,R) increases by w(L,v);
- w(L, M) decreases by w(L, v);
- w(M,R) increases by w(M,v) and decreases by w(v,R);
- w(M, M) decreases by w(v, M) and decreases by w(M, v).

Overall, Φ increases by

$$4w(L,v) - 2w(L,v) + 2(w(M,v) - w(v,R)) - (w(v,M) + w(M,v))$$

$$= 2w(L,v) + w(M,v) - 2w(v,R) - w(v,M),$$

and this quantity is positive when the algorithm decides to move v to R. Upon the end of the algorithm, we have $\Phi = 4w(L,R)$ due to $M = \emptyset$, whereas, initially, $\Phi = w(M,M) = w(V,V)$ due to M = V. Since Φ is nondecreasing, we conclude that $4w(L,R) \ge w(V,V)$ holds as claimed.

The following lemma and its corollary are needed to evaluate the $w_{k,h}$ functions of Construction 5.7.

Lemma 5.20. Given a text $T \in \Sigma^n$, represented by a straight-line grammar of size g, and the interval representation of a set $P \subseteq [1..n]$, in $\mathcal{O}((g+|\mathcal{I}(P)|)\log n)$ time one can compute, for every length-1 substring a of T, the value $|\{i \in P : T[i] = a\}|$.

Proof. First, we construct an AVL grammar \mathcal{G} of size $\mathcal{O}(g \log n)$ that generates T; see [Ryt03]. Next, for every interval I = [b ... e) in the representation of P, we extend the grammar with a symbol X_I whose expansion is T[b ... e). Finally, we add a new start symbol to \mathcal{G} whose production is $\bigcirc_I X_I$, where I iterates over the intervals representing P. This grammar can be constructed in $\mathcal{O}((g + |\mathcal{I}(\mathsf{P})|) \log n)$ time.

It is easy to see that, for every length-1 substring a of T, the value $|\{i \in P : T[i] = a\}|$ equals the number of occurrences of a in the string produced by \mathcal{G} . This corresponds to the number of paths from the start

symbol to a in the DAG representing \mathcal{G} ; such values can be computed in $\mathcal{O}(|\mathcal{G}|) = \mathcal{O}((g + |\mathcal{I}(\mathsf{P})|) \log n)$ time.

Corollary 5.21. Given a text $T \in \Sigma^n$, represented by an SLP of size g, and the interval representation of a set $P \subseteq [1..n)$, in $\mathcal{O}((g + |\mathcal{I}(P)|) \log n)$ time one can compute, for every length-2 substring ab of T, the value $|\{i \in P : T[i..i+1] = ab\}|$.

Proof. We construct a text $\hat{T} \in (\Sigma^2)^{n-1}$ such that $\hat{T}[i] = T[i \dots i+1]$. For this, we transform the input grammar \mathcal{G} into a grammar $\hat{\mathcal{G}}$. For each symbol A with $\mathsf{rhs}_{\mathcal{G}}(A) = BC$, we create a symbol \hat{A} with $\mathsf{rhs}_{\hat{\mathcal{G}}}(\hat{A}) = \hat{B} \cdot (\mathsf{RS}(B), \mathsf{LS}(C)) \cdot \hat{C}$. If the starting symbol of \mathcal{G} is S, then the starting symbol of $\hat{\mathcal{G}}$ is \hat{S} . It is easy to see that $\exp(\hat{S}) = \hat{T}$ and that $\{i \in \mathsf{P} : T[i \dots i+1] = ab\} = \{i \in \mathsf{P} : \hat{T}[i] = (a,b)\}$. Thus, it suffices to use Lemma 5.20.

We are now ready to provide the implementation of the construction algorithm.

Theorem 5.22. Suppose that we are given the LZ77 parsing of a text $T \in [0..\sigma)^n$ and, for every h with $m_h \leq n$, the interval representation $\mathcal{I}(\mathsf{C}_{m_h})$ of size $\mathcal{O}(\frac{1}{m_h}|\mathsf{C}_{m_h}|)$. Then, the RLSLP of Constructions 5.6 and 5.7 can be constructed in $\mathcal{O}(C\log^6 n)$ time, where $C = \sum_{h=0}^{\infty} \frac{1}{m_h}|\mathsf{C}_{m_h}|$.

Proof. Our main goal is to build strings $(T_k)_{k=0}^{\infty}$ for subsequent integers $k \in \mathbb{Z}_{\geq 0}$. For each iteration, the output consists of the set $\mathcal{S}_{\leq k} := \bigcup_{h=0}^{k} \mathcal{S}_h$ as well as the LZ77 parsing of T_k , where each symbol is stored as a pointer to an element of $\mathcal{S}_{\leq k}$. Moreover, for every non-terminal $A \in \mathcal{S}_{\leq k}$, if $A = (B, C) \in \mathcal{A}^2$, then $B, C \in \mathcal{S}_{\leq k}$ are represented as pointers to elements of $\mathcal{S}_{\leq k}$. Similarly, if $A = (B, m) \in \mathcal{S}_{\leq k} \cap \mathcal{A} \times \mathbb{Z}_{\geq 2}$, then B is represented as a pointer to an element of $\mathcal{S}_{\leq k}$. Each symbol $A \in \mathcal{S}_{\leq k}$ also stores $|\exp(A)|$, i.e., the length of its expansion.

In the base case of k = 0, the string $T_0 = T$ is already represented by its LZ77-parsing. A linear scan over this parsing lets us retrieve the set $S_{\leq 0}$ of symbols that occur in T and alter the parsing so that symbols are stored using pointers to $S_{\leq 0}$.

Once the algorithm reaches $|T_k| = 1$, which happens for $k = \mathcal{O}(\log n)$ by Corollary 5.16, the set $\mathcal{S}_{\leq k}$ forms an RLSLP \mathcal{G} generating T. Moreover, Theorem 5.17 guarantees that $|\mathcal{S}_{\leq k}| = |\mathcal{S}| = \mathcal{O}(C)$.

By Fact 5.8, if we change \mathcal{G} so that all symbols in $\mathcal{S}_{\leq h}$ become terminals, then \mathcal{G} would generate T_h instead. Consequently, $z(T_h) = \mathcal{O}(C)$ holds for all $h \in \mathbb{Z}_{>0}$.

It remains to implement the iteration, where we derive T_k from T_{k-1} for $k \in \mathbb{Z}_{>0}$. If k is odd, then our algorithm applies Lemma 5.3 with $\mathcal{B} = \mathcal{A}_k \cap \mathcal{S}_{k-1}$. This subroutine costs $\mathcal{O}(z(T_{k-1}) \cdot \log^5 n) = \mathcal{O}(C \log^5 n)$ time and produces the LZ77 representation of T_k . A left-to-right scan over this representation lets us determine $\mathcal{S}_k \setminus \mathcal{S}_{k-1}$ (which is equal to $\mathcal{S}_{\leq k} \setminus \mathcal{S}_{\leq (k-1)}$ by Fact 5.8). For each symbol in $A \in \mathcal{S}_k \setminus \mathcal{S}_{k-1}$, we add an appropriate entry to $\mathcal{S}_{\leq k}$ and replace its representation (of the form $(B, m) \in \mathcal{S}_{\leq k-1} \times \mathbb{Z}_{\geq 2}$) with a pointer to $\mathcal{S}_{\leq k}$.

The algorithm for even k is more difficult because we first need to find sets \mathcal{L}_k , $\mathcal{R}_k \subseteq \mathcal{A}_k \cap \mathcal{S}_{k-1}$ satisfying Construction 5.7, that is, $w_k(\mathcal{L}_k, \mathcal{R}_k) \geq \frac{1}{4} w_k(\mathcal{A}_k, \mathcal{A}_k)$. For this, we convert the LZ77-representation of T_{k-1} into an AVL grammar \mathcal{G}_k using Theorem 5.1; the size of this grammar is $\mathcal{O}(z(T_{k-1})\log n) \subseteq \mathcal{O}(C\log n)$ and the grammar is constructed in $\mathcal{O}(C\log n)$ time. For each symbol $X \in T_{\mathcal{G}_k}$, we compute the value $|\exp(\exp_{\mathcal{G}_k}(X))|$. Since \mathcal{G}_k is of logarithmic height, given any $i \in [0 \dots n]$, we can compute the rank of i in B_{k-1} in $\mathcal{O}(\log n)$ time. In particular, for every $h \in \mathbb{Z}_{\geq 0}$, this lets us convert the interval representation of C_{m_k} into the interval representation of $J_{k,h}$ of size $|\mathcal{I}(J_{k,h})| = |\mathcal{I}(C_{m_k})| = \mathcal{O}(\frac{1}{m_k}|C_{m_k}|) = \mathcal{O}(C)$. We execute this step for all $h \in [k \dots k]$, where $\kappa = 2\lceil \log_{8/7} n \rceil$ is defined in Lemma 5.15 and satisfies $m_{\kappa} \geq n$. Now, we can use Corollary 5.21 to compute all the non-zero values of the $w_{k,h}$ function. Since $|\mathcal{G}_k| = \mathcal{O}(C\log n)$, this step costs $\mathcal{O}(C\log n + C\log^2) = \mathcal{O}(C\log^2 n)$ time for each $h \in [0 \dots \kappa]$ and $\mathcal{O}(C\log^3 n)$ time in total. To retrieve $w_k(A, B)$, we note that $w_{k,h}(A, B) = w_{k,\kappa}(A, B)$ holds for $h \geq \kappa$ due to $C_{m_k} = [1 \dots n] = C_{m_\kappa}$. Hence, since κ is even,

$$\sum_{h=k}^{\infty} \left(\frac{3}{4}\right)^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} w_{k,h}(A,B) = \sum_{h=k}^{\kappa-1} \left(\frac{3}{4}\right)^{\lfloor \frac{h}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} w_{k,h}(A,B) + 8 \cdot \left(\frac{3}{4}\right)^{\lfloor \frac{\kappa}{2} \rfloor - \lfloor \frac{k}{2} \rfloor} w_{k,\kappa}(A,B).$$

Consequently, we can retrieve $w_k(A, B)$ by adding up the values $w_{k,h}(A, B)$ for $h \in [k ... \kappa]$ with appropriate coefficients (which are integers when scaled up by a factor of $2^{\kappa} = n^{\mathcal{O}(1)}$). Thus, it takes $\mathcal{O}(C \log^3 n)$ time to construct a multigraph whose vertices are $\mathcal{B}_k \cap \mathcal{S}_{k-1}$ and with edges (A, B) with (integer) weights $2^{\kappa} \cdot w_k(A, B)$ (we do not add the edge if AB is not a substring of T_{k-1} ; in that case, its weight would be 0 anyway). By Corollary 5.9, this graph does not contain loops. Therefore, we can use Lemma 5.19 to obtain the desired sets $\mathcal{L}_k, \mathcal{R}_k \subseteq \mathcal{A}_k \cap \mathcal{B}_{k-1}$ satisfying $w_k(\mathcal{L}_k, \mathcal{R}_k) \geq \frac{1}{4} w_k(\mathcal{A}_k, \mathcal{A}_k)$. The overall cost of this subroutine is $\mathcal{O}(C \log^3 n)$. Finally, our algorithm applies Lemma 5.5. This subroutine costs $\mathcal{O}(z(T_{k-1}) \cdot \log^4 n) = \mathcal{O}(C \log^5 n)$ time

and produces the LZ77 representation of T_k . A left-to-right scan over this representation lets us determine $\mathcal{S}_k \setminus \mathcal{S}_{k-1}$ (which is equal to $\mathcal{S}_{\leq k} \setminus \mathcal{S}_{\leq (k-1)}$ by Fact 5.8). For each symbol in $A \in \mathcal{S}_k \setminus \mathcal{S}_{k-1}$, we add an appropriate entry to $S_{\leq k}$ and replace its representation (of the form $(B, C) \in S_{\leq k-1}^2$) with a pointer to $S_{\leq k}$. The total cost of the algorithm (across all iterations) is $\mathcal{O}(C \log^6 n)$ due to Corollary 5.16.

Corollary 5.23. Given the LZ77 parsing of a text $T \in [0...\sigma)^n$, the RLSLP of Constructions 5.6 and 5.7 can be constructed in $\mathcal{O}(\delta(T) \cdot \log^7 n)$ time assuming that the cover hierarchy is chosen according to Construction 4.6. Moreover, the size of this RLSLP is $\mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$.

Proof. First, we use Corollary 4.9 to compute $\mathcal{I}(\mathsf{C}_{m_h})$ for all $h \in \mathbb{Z}_{\geq 0}$ such that $m_h \leq n$; this step costs $\mathcal{O}(z(T)\log^4 n) = \mathcal{O}(\delta(T)\log^5 n)$ time for each h and $\mathcal{O}(\delta(T)\log^6 n)$ time in total. Next, we use the algorithm of Theorem 5.22. Note that $C = \mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ holds due to Lemma 5.18, so the total running time is $\mathcal{O}(C\log^6 n) = \mathcal{O}(\delta(T)\log^7 n).$

5.4 Consequences

Optimal Compressed Space Random Access

Theorem 5.24. For any text $T \in [0..\sigma)^n$, there exists a data structure of size $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ that, given any $i \in [1..n]$, returns T[i] in $\mathcal{O}(\log n)$ time. Moreover, given the LZ77-parsing of T, it can be constructed in $\mathcal{O}(\delta(T)\log^7 n)$ time.

Proof. Our data structure consists of the RLSLP constructed in $\mathcal{O}(\delta(T)\log^7 n)$ time using Corollary 5.23. At the query time, we traverse the parse tree of T from the root to the leaf representing T[i]. The query time is proportional to the RLSLP height, which is $\mathcal{O}(\log n)$ by Corollary 5.16.

Optimal Compressed Space LCE Queries

Theorem 5.25. For any text $T \in [0..\sigma)^n$, there exists a data structure of size $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ answering LCE_T and LCE_T queries in $\mathcal{O}(\log n)$ time. Moreover, given the LZ77-parsing of T, it can be constructed in $\mathcal{O}(\delta(T)\log^7 n)$ time.

Proof. Our data structure consists of the RLSLP constructed in $\mathcal{O}(\delta(T)\log^7 n)$ time using Corollary 5.23. In the following, we describe an $\mathcal{O}(\log n)$ -time algorithm answering an LCE_T(i,i') query; the procedure for an $LCE_{\overline{T}}$ query is symmetric. We assume $i \neq i'$ without loss of generality; otherwise, $LCE_T(i,i) = n - i + 1$ can be computed trivially.

The query procedure maintains two nodes ν, ν' of the parse tree \mathcal{T} whose expansions $\exp(\nu) = T[\ell ... r]$ and $\exp(\nu') = T[\ell' ... r')$ satisfy $T[i...\ell] = T[i'...\ell']$. Technically, each of these nodes is associated with a stack storing (the expansions and symbols of) all its ancestors.

We initialize ν as the highest node such that $\exp(\nu)$ starts at position i, and ν' as the highest node such that $\exp(\nu')$ starts at position i'. For this, we follow the paths from the root to the leaves representing T[i]and T[i'], respectively. At each step, we proceed as follows:

- 1. If $|\exp(\nu)| = |\exp(\nu')| = 1$ and $s(\nu) \neq s(\nu')$, return ℓi .
- 2. If $|\exp(\nu)| > |\exp(\nu')|$, we replace ν by its leftmost child.
- 3. If $|\exp(\nu')| > |\exp(\nu)|$, we replace ν' by its leftmost child.
- 4. If $s(\nu) \neq s(\nu')$, we replace both ν and ν' by their leftmost children.

- 5. Suppose that ν is the jth among the d children of its parent and ν' is the j'th among the d' children of its parent. Let $\lambda = \min(d j, d' j')$. If $\lambda \leq 1$, we replace ν and ν' the highest nodes whose expansions start at position r and r', respectively.
- 6. Otherwise, we replace ν by its $(j + \lambda)$ th sibling and ν' by its $(j' + \lambda)$ th sibling.

Let us justify the correctness of the algorithm. In case 1, we have $T[i \dots \ell) = T[i' \dots \ell']$ yet $T[\ell] = \mathsf{s}(v) \neq \mathsf{s}(\nu') = T[\ell']$. Hence, $\mathrm{LCE}_T(i,i') = \ell - i = \ell' - i'$ is computed correctly. In cases 2–4, the invariant is still satisfied because the leftmost positions in $\exp(\nu)$ and $\exp(\nu')$ do not change. Moreover, whenever we replace ν or ν' with its leftmost child, we have $|\exp(\nu)| > 1$ and $|\exp(\nu')| > 1$, respectively, so the child exists. In the remaining cases, we have $\mathsf{s}(\nu) = \mathsf{s}(\nu')$ and thus $T[i \dots r) = T[i' \dots r']$. Thus, the invariant still holds in case 5, where we replace ν and ν' by the highest nodes whose expansions start at positions r and r', respectively. If $\lambda > 1$, then $d, d' \geq 3$, and thus all siblings of ν and ν' have the same symbol. Thus, the invariant holds because we shift ν and ν' by λ siblings to the right.

It remains to analyze the query time. For this, we say that nodes (ν, ν') form a matching pair if $s(\nu) = s(\nu')$ and the expansions $\exp(\nu) = T[\ell \dots r)$ and $\exp(\nu) = T[\ell \dots r)$ satisfy $T[i \dots \ell] = T[i' \dots \ell']$. Denote by N and N' the set of nodes ν and ν' participating in a matching pair; these pairs form a perfect matching between N and N'. We claim that the query algorithm satisfies the following additional invariant: None of the ancestors of ν belongs to N and none of the ancestors of ν' belong to N'. The invariant is clearly satisfied at the beginning because all the ancestors of ν and ν have their expansions starting before position i and i', respectively. In cases 2 and 4, the node ν does for a matching pair with any ancestor of ν' (because they belong to N'), with the node ν itself (because $s(\nu) \neq s(\nu')$), nor with any descendant of ν' (because their expansions are shorter than $\exp(\nu)$). Thus, ν does not belong to N. By symmetry, the same holds in cases 3 and 4 when we replace ν' by its leftmost child. In cases 5 and 6, all the ancestors of the new nodes ν and ν' are also the ancestors of the old nodes ν and ν' .

Let $\hat{N} \subseteq N$ consist of nodes of N whose ancestors do not belong to N, and define \hat{N}' analogously. Our algorithm visits only nodes in \hat{N} and \hat{N}' and their ancestors, as well as the ancestors of the two leaves whose mismatch causes the algorithm to return. Moreover, if \hat{N} or \hat{N}' contains three or more siblings, then Case 6 guarantees that we visit at most two of them. We claim that \hat{N} and \hat{N}' have $\mathcal{O}(\log n)$ parents in total. Specifically, we claim that each level of the parsing contributes nodes with at most two parents. For this, let us fix a level and let $T_k[j]$ and $T_k[j]$ be the leftmost and the rightmost symbol of T_k that corresponds to a node in \hat{N} . By definition of matching pairs, there are analogous symbols $T_k[j']$ and $T_k[j']$ corresponding to nodes on \hat{N}' . By Fact 5.8, the fragments $T_k[j \dots j]$ and $T_k[j' \dots j']$ match. The block boundaries of Construction 5.6 depend only on the adjacent two symbols, so the block boundaries strictly within $T_k[j \dots j]$ and $T_k[j' \dots j']$ are placed analogously. Hence, the only symbols of $T_k[j \dots j]$ corresponding to nodes in \hat{N} might be located before the leftmost block boundary within $T_k[j \dots j]$ or after the rightmost block boundary within $T_k[j \dots j]$. These symbols belong to two blocks, and thus the corresponding nodes have two parents. By Corollary 5.16, the parsing has $\mathcal{O}(\log n)$ levels, so the total query time is $\mathcal{O}(\log n)$.

5.4.3 Optimal Compressed Space Synchronizing Sets

Definition 5.26 $(\tau\text{-runs})$. For a string $T \in \Sigma^+$ and an integer $\tau \in [1 ... n]$, we define the set $\mathsf{RUNS}_{\tau}(T)$ of τ -runs in T that consists of all fragments T[p ... q] of length at least τ that satisfy $\mathsf{per}(T[p ... q]) \leq \frac{1}{3}\tau$ yet cannot be extended (in any direction) while preserving the shortest period.

Construction 5.27 ([KRRW23]). For $\tau \in [1 .. \lfloor \frac{n}{2} \rfloor]$, let $k = \max\{h \in \mathbb{Z}_{\geq 0} : \tau \geq \alpha_h\}$. A position $i \in [1 .. n - 2\tau + 1]$ is included in $S(\tau, T)$ if at least one of the following conditions holds:

- (1) $i + \tau 1 \in \mathsf{B}_k$ and $T[i ... i + 2\tau)$ is not contained in any τ -run;
- (2) i = p 1 for some τ -run $T[p ... q] \in \mathsf{RUNS}_{\tau}(T)$;
- (3) $i = q 2\tau + 2$ for some τ -run $T[p ... q] \in \mathsf{RUNS}_{\tau}(T)$.

Lemma 5.28. The set $S(\tau,T)$ obtained using Construction 5.27 is a τ -synchronizing set. Moreover, $|\{i \in S(\tau,T) : i+\tau-1 \in C_{m_k}\}| = \mathcal{O}(|\mathsf{B}_k \cap \mathsf{C}_{m_k}| + \frac{1}{m_k}|\mathsf{C}_{m_k}|)$, where $k = \max\left(\{0\} \cup \{h \in \mathbb{Z}_{>0} : \tau \geq \alpha_h\}\right)$.

Proof. First, suppose that $i, i' \in [1 \dots n - 2\tau + 1]$ satisfy $T[i \dots i + 2\tau) = T[i' \dots i' + 2\tau)$ and $i \in S(\tau, T)$. We will show that if i satisfies one of the conditions (1)–(3), then i' satisfies the same condition. If i satisfies condition (1), then $\operatorname{per}(T[i' \dots i' + 2\tau)) = \operatorname{per}(T[i \dots i + 2\tau)) > \frac{1}{3}\tau$, so $T[i' \dots i' + 2\tau)$ is not contained in any τ -run. At the same time, due to $T(i+\tau-1-\alpha_k\dots i+\tau-1+2\alpha_k] = T(i'+\tau-1-\alpha_k\dots i'+\tau-1+2\alpha_k]$, by Lemma 5.11, $i+\tau-1 \in \mathsf{B}_k$ implies $i'+\tau-1 \in \mathsf{B}_k$. Consequently, i' also satisfies condition (1). If i satisfies condition (2), then $\operatorname{per}(T[i+1\dots i+\tau]) = \operatorname{per}(T[i'+1\dots i'+\tau]) \le \frac{1}{3}\tau < \operatorname{per}(T[i\dots i+\tau]) = \operatorname{per}(T[i'\dots i'+\tau])$. Hence, $T[i'+1\dots i'+\tau]$ can be extended to a τ -run $T[p'\dots q']$ that starts at position p'=i'+1, and thus i' satisfies condition (2). Similarly, if i satisfies condition (3), then $\operatorname{per}(T[i'+\tau-1\dots i'+2\tau-2]) = \operatorname{per}(T[i+\tau-1\dots i+2\tau-1]) \le \frac{1}{3}\tau < \operatorname{per}(T[i+\tau-1\dots i'+2\tau-1])$. Hence, $T[i'+\tau-1\dots i'+2\tau-1]$ can be extended to a τ -run $T[p'\dots q']$ that ends at position $q'=i'+2\tau-2$, and thus i' satisfies condition (3).

For a proof of the density condition, consider a position $i \in [1 ... n - 3\tau + 2]$ with $[i ... i + \tau) \cap S(\tau, T) = \emptyset$. We start by identifying a τ -run T[p ... q] with $p \le i + \tau$ and $q \ge i + 2\tau - 2$. First, suppose that there exists a position $b \in [i + \tau - 1 ... i + 2\tau - 1) \cap B_k$. Since $b - \tau + 1 \in [i ... i + \tau)$ has not been added to $S(\tau, T)$, the fragment $T[b - \tau + 1 ... b + \tau]$ must be contained in a τ -run that satisfies $p \le b - \tau \le i + \tau - 1$ and $q \ge b + \tau \ge i + 2\tau - 1$.

Next, suppose that $[i+\tau \ldots i+2\tau)\cap \mathsf{B}_k=\emptyset$. Then, $T[i+\tau-1\ldots i+2\tau]$ is contained in a single phrase induced by T_k , and the length of this phrase is at least $\tau+1$. If k=0, this contradicts the fact that all level-0 phrases are of length 1. Otherwise, $\tau+1\geq\alpha_k+1>2\ell_k$, so Fact 5.10 yields $\operatorname{per}(T[i+\tau-1\ldots i+2\tau-1])\leq\ell_k\leq\frac13\tau$. The τ -run $T[p\ldots q]$ extending $T[i+\tau-1\ldots i+2\tau]$ satisfies $p\leq i+\tau-1$ and $q\geq i+2\tau$.

Note that positions i=p-1 and $i=q-2\tau+2$ satisfy conditions (2) and (3), respectively. Due to $[i ... i+\tau) \cap \mathsf{S}(\tau,T) = \emptyset$, this implies $p \leq i$ and $q \geq i+3\tau-2$, which means that $\mathrm{per}(T[i ... i+3\tau-1)) \leq \frac{1}{3}\tau$ holds as claimed.

For the converse implication, note that if $s \in [i ... i + \tau) \cap S(\tau, T)$, then $\operatorname{per}(T[i ... i + 3\tau - 1)) \geq \operatorname{per}(T[s ... s + 2\tau)) > \frac{1}{3}\tau$ because $T[s ... s + 2\tau)$ is not contained in any τ -run (the latter observation is trivial if s satisfies condition (1); in the remaining two cases, it follows from the upper bound of $\frac{2}{3}\tau$ on the overlap of two τ -runs).

As for the size of $S(\tau, T)$, observe that if $i \in S(\tau, T)$, then $i + \tau - 1 \in B_k$ or there is a τ -run T[p ...q] with p = i + 1 or $q = i + 2\tau - 2$. The contribution of the first term is $|B_k \cap C_{m_k}|$. The latter two cases contribute $\mathcal{O}(|\mathcal{I}(C_{m_k})| + \frac{1}{m_k}|C_{m_k}|) = \mathcal{O}(\frac{1}{m_k}|C_{m_k}|)$ due to the upper bound of $\frac{2}{3}\tau$ on the overlap of two τ -runs. \square

Lemma 5.29. Given the LZ77 representation of a text $T \in [0...\sigma)^n$, and integer $\tau \in [1...\lfloor \frac{n}{2}\rfloor]$, and the interval representation of the set C_{m_k} such that $k = \max(\{0\} \cup \{h \in \mathbb{Z}_{>0} : \tau \geq \alpha_h\})$, the set $\{i \in S(\tau, T) : i + \tau - 1 \in C_{m_k}\}$ can be constructed in $\mathcal{O}(\delta(T) \cdot \log^7 n + (|B_k \cap C_{m_k}| + \frac{1}{m_k}|C_{m_k}|) \log^3 n)$ time.

Proof. We use the LCE queries of Theorem 5.25 as well as the period queries of [KK20b, Theorem 6.7]: Given any fragment x of T, in $\mathcal{O}(\log^3 n)$ time, we can compute $\operatorname{per}(x)$ or report that $\operatorname{per}(x) > \frac{1}{2}|x|$. The total preprocessing time is $\mathcal{O}(\delta(T) \cdot \log^7 n)$.

We separately process each interval I in the interval representation of the set C_{m_k} . Let $I = [\ell \dots r]$. Our first goal is to compute all runs in $\mathsf{RUNS}_\tau(T)$ that have at least τ characters in common with $T[\ell \dots r+2\tau)$. We partition $T[\ell \dots r+2\tau)$ into blocks of length $\lfloor \frac{1}{3}\tau \rfloor$ (leaving up to $\lfloor \frac{1}{3}\tau \rfloor$ trailing characters behind). For any two consecutive blocks, we apply a 2-period query to determine the shortest period of their concatenation (provided that it does not exceed $\lfloor \frac{1}{3}\tau \rfloor$). If the period does not exceed $\lfloor \frac{1}{3}\tau \rfloor$, we use LCE queries of Theorem 5.25 to maximally extend the fragment while preserving its shortest period. If the extended fragment has at least τ characters in common with $T[\ell \dots r+2\tau)$, we save it as one of the desired τ -runs. For each of the computed runs, we report the appropriate positions on the output. This step costs $\mathcal{O}(\lceil |I|/\tau \rceil \log^3 n)$ time per interval and $\mathcal{O}((|\mathcal{I}(\mathsf{C}_{m_k})| + \frac{1}{m_k}|\mathsf{C}_{m_k}|)\log^3 n) = \mathcal{O}(\frac{1}{m_k}|\mathsf{C}_{m_k}|\log^3 n)$ time in total.

Next, we construct the set $\{i: i+\tau-1\in\mathsf{B}_k\cap\mathsf{C}_{m_k}\}$ by traversing the parse tree of T. For each position

Next, we construct the set $\{i: i+\tau-1 \in \mathsf{B}_k \cap \mathsf{C}_{m_k}\}$ by traversing the parse tree of T. For each position i in this set, we use a period query to check if $T[i...i+2\tau)$ is contained in any run and, if not, add i to the output. This step costs $\mathcal{O}(|\mathsf{B}_k \cap \mathsf{C}_{m_k}|\log^3 n)$ time.

Proposition 5.30. Let $T \in [0..\sigma)^n$ and $c \in \mathbb{Z}_{>0}$. Given the LZ77 parsing of T, in $\mathcal{O}(c \cdot \delta(T) \log^7 n)$ time we can construct a collection $\{S_{\text{comp},i}\}_{i \in [4..\lceil \log n \rceil)}$ such that:

- For every $i \in [4..\lceil \log n \rceil)$, letting $\tau_i = \lfloor \frac{2^i}{3} \rfloor$, it holds $\mathsf{S}_{\mathrm{comp},i} = \mathrm{comp}_{c\tau_i}(\mathsf{S}_i,T) = \mathsf{S}_i \cap \mathsf{C}(c\tau_i,T)$ (Definition 4.10), where S_i is a τ_i -synchronizing set of T and $\mathsf{C}(c\tau_i,T)$ is as in Construction 4.6,
- It holds $\sum_{i \in [4..\lceil \log n \rceil)} |\mathsf{S}_{\text{comp},i}| = \mathcal{O}(c\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n}).$

Proof. For $\ell \in \mathbb{Z}_{>0}$, define $\mathsf{C}_{\ell} := \bigcup_{i \in \mathsf{C}(c\ell,T)} [i \ldots i + \ell) \cap [1 \ldots n]$. We shall prove that $(\mathsf{C}_{\ell})_{\ell \in \mathbb{Z}_{>0}}$ forms a cover hierarchy of T. Since $\mathsf{C}(\ell,T) \subseteq \mathsf{C}(c\ell,T) \subseteq \mathsf{C}_{\ell}$, the set C_{ℓ} is an ℓ -cover. Moreover, $|\mathcal{I}(\mathsf{C}_{\ell})| \leq \mathsf{C}_{\ell}$ $|\mathcal{I}(\mathsf{C}(c\ell,T))| \leq \max(1,\frac{1}{c\ell}|\mathsf{C}(c\ell,T)|) \leq \max(1,\frac{1}{c\ell}|\mathsf{C}_{\ell}|) \leq \max(1,\frac{1}{\ell}|\mathsf{C}_{\ell}|).$ It remains to prove that the family C_{ℓ} is monotone. For this, consider $\ell, \ell' \in \mathbb{Z}_{>0}$ with $\ell \leq \ell'$ and a position $j \in C_{\ell}$. By definition of C_{ℓ} , there exists $i \in \mathsf{C}(c\ell,T)$ such that $j \in [i \dots i + \ell)$. Since $\mathsf{C}(c\ell,T) \subseteq \mathsf{C}(c\ell',T)$ and $[i \dots i + \ell) \subseteq [i \dots i + \ell')$, we conclude that $j \in C_{\ell'}$ holds as claimed. Observe now that by a simple generalization of Lemma 5.18, it holds $\sum_{k=0}^{\infty} \frac{1}{m_k} |\mathsf{C}_{m_k}| = \mathcal{O}(c\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n}).$

The algorithm proceeds as follows:

- Note that, for each $\ell \in \mathbb{Z}_{>0}$, the interval representation $\mathcal{I}(\mathsf{C}_{\ell})$ can be constructed in $\mathcal{O}(z(T)\log^4 n)$ time using Corollary 4.9 to build $\mathcal{I}(\mathsf{C}(c\ell,T))$. We apply Theorem 5.22 on top of this construction. By the above, this takes $\mathcal{O}(c \cdot \delta(T) \log^7 n)$ time.
- We preprocess the LZ77 parsing of T using Lemma 5.29 in $\mathcal{O}(\delta(T)\log^7 n)$ time. Consider $i \in$ $[4..\lceil \log n \rceil)$ and let $k = \max(\{0\} \cup \{h \in \mathbb{Z}_{>0} : \tau_i \geq \alpha_h\})$. First, using Lemma 5.29, in $\mathcal{O}(|\mathsf{B}_k \cap \mathsf{B}_k|)$ $C_{m_k}|+\frac{1}{m_k}|C_{m_k}|\log^3 n$) time we compute the set $\{i\in S(\tau_i,T):i+\tau-1\in C_{m_k}\}$. By definition of $C_{\tau_i},j\in C(c\tau_i,T)$ implies $[j...j+\tau_i)\subseteq C_{\tau_i}\subseteq C_{m_k}$ (where $C_{\tau_i}\subseteq C_{m_k}$ holds since $\tau_i\leq \alpha_{k+1}\leq m_k$). In particular, $j \in C(c\tau_i, T)$ implies $j + \tau - 1 \in C_{m_k}$. Consequently, $S(\tau_i, T) \cap C(c\tau_i, T) \subseteq \{i \in S(\tau_i, T) : t \in S(\tau_i, T) : t \in S(\tau_i, T) \}$ $i+\tau-1\in \mathsf{C}_{m_k}$. Using Corollary 4.9, in $\mathcal{O}(z(T)\log^4 n)$ we thus construct $\mathcal{I}(\mathsf{C}(c\tau_i,T))$ and filter the set $\{i \in \mathsf{S}(\tau_i,T) : i+\tau-1 \in \mathsf{C}_{m_k}\}\$ to obtain $\mathsf{S}(\tau_i,T) \cap \mathsf{C}(c\tau_i,T)$. By $\sum_{k=0}^{\infty} \frac{1}{m_k} |\mathsf{C}_{m_k}| = \mathcal{O}(c\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ and the proof of Theorem 5.17, the application of Lemma 5.29 over all i takes $\mathcal{O}(c\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n}\log^3 n)$ time.

In total, we thus spend $\mathcal{O}(c\delta(T)\log^7 n)$ time. By Lemma 5.28, the summation in the proof of Theorem 5.17, and the upper bound $\sum_{k=0}^{\infty} \frac{1}{m_k} |\mathsf{C}_{m_k}| = \mathcal{O}(c\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$, the total size of the sets $\mathsf{S}(\tau_i, T) \cap \mathsf{C}(c\tau_i, T)$ is $\mathcal{O}(c\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n}).$

Weighted Range Queries 6

Let \mathcal{X} and \mathcal{Y} be some linearly ordered sets (we denote the order on both sets using \prec or \preceq). Let $\mathcal{P} \subseteq$ $\mathcal{X} \times \mathcal{Y} \times \mathbb{Z}_{>0} \times \mathbb{Z}$ be a finite set of points, where each point is associated with some positive integer weight and some integer label. Unless explicitly stated otherwise, we do not place any restriction on \mathcal{P} , and only require that \mathcal{P} is a set (not a multiset). In particular, \mathcal{P} can in general contain two points with equal coordinates as long as they differ either on the weight or on the label.

Weighted range counting: Let $x_l, x_u \in \mathcal{X}$ and $y_l, y_u \in \mathcal{Y}$. We define

- $$\begin{split} \bullet & \text{ weight-count}_{\mathcal{P}}(x_l, x_u, y_l, y_u) \coloneqq \sum_{(x, y, w, \ell) \in \mathcal{R}_4} w, \\ \bullet & \text{ weight-count}_{\mathcal{P}}(x_l, x_u, y_u) \coloneqq \sum_{(x, y, w, \ell) \in \mathcal{R}} w, \\ \bullet & \text{ weight-count}_{\mathcal{P}}^{\preceq}(x_l, x_u, y_u) \coloneqq \sum_{(x, y, w, \ell) \in \mathcal{R}_3^{\preceq}} w, \\ \bullet & \text{ weight-count}_{\mathcal{P}}(x_l, x_u) \coloneqq \sum_{(x, y, w, \ell) \in \mathcal{R}_2} w, \end{aligned}$$

- $$\begin{split} \bullet & \ \mathcal{R}_4 = \{(x,y,w,\ell) \in \mathcal{P} : x_l \preceq x \prec x \text{ and } y_l \preceq y \prec y\}, \\ \bullet & \ \mathcal{R}_3 = \{(x,y,w,\ell) \in \mathcal{P} : x_l \preceq x \prec x_u \text{ and } y \prec y_u\}, \\ \bullet & \ \mathcal{R}_3^{\preceq} = \{(x,y,w,\ell) \in \mathcal{P} : x_l \preceq x \prec x_u \text{ and } y \preceq y_u\}, \\ \bullet & \ \mathcal{R}_2 = \{(x,y,w,\ell) \in \mathcal{P} : x_l \preceq x \prec x_u\}. \end{split}$$

Weighted range selection: Let $x_l, x_u \in \mathcal{X}$ and $r \in [1...\text{weight-count}_{\mathcal{P}}(x_l, x_u)]$. We define

weight-select_{$$\mathcal{P}$$} $(x_l, x_u, r) := \{ \ell \in \mathbb{Z} : (x, y, w, \ell) \in \mathcal{P}, x_l \leq x \prec x_u, \text{ and } y = y_u \},$

where $y_u \in \mathcal{Y}$ is such that $r \in (\text{weight-count}_{\mathcal{P}}(x_l, x_u, y_u))$. weight-count $\mathcal{P}(x_l, x_u, y_u)$. A weighted range selection query asks to return any element of the set weight-select $\mathcal{P}(x_l, x_u, r)$.

Range minimum: Let $x_l, x_u \in \mathcal{X}$ and $y_l, y_u \in \mathcal{Y}$ be such that weight-count_{\mathcal{P}} $(x_l, x_u, y_l, y_u) > 0$. We define

- $\operatorname{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_u) := \min_{(x, y, w, \ell) \in \mathcal{R}_4} \ell$,
- $\operatorname{r-min}_{\mathcal{P}}(x_l, x_u) := \min_{(x, y, w, \ell) \in \mathcal{R}_2} \ell,$

where

• $\mathcal{R}_4 = \{(x, y, w, \ell) \in \mathcal{P} : x_l \leq x \prec x_u \text{ and } y_l \leq y \prec y_u\},$ • $\mathcal{R}_2 = \{(x, y, w, \ell) \in \mathcal{P} : x_l \leq x \prec x_u\}.$

6.1 Integer-Integer Coordinates

Proposition 6.1. Let $\epsilon > 0$ be a fixed constant. Given a set \mathcal{P} of m points with coordinates in $\mathbb{Z}_{\geq 0}$, such that for every $(x, y, w, \ell), (x', y', w', \ell') \in \mathcal{P}$, (x, y) = (x', y') implies w' = w and $\ell' = \ell$, we can in $\mathcal{O}(m \log m)$ time construct a structure of size $\mathcal{O}(m)$ supporting the following queries on \mathcal{P} :

- Weighted range counting queries in $\mathcal{O}(\log^{2+\epsilon} m)$ time,
- Weighted range selection queries in $\mathcal{O}(\log^{3+\epsilon} m)$ time,
- Range minimum queries in $\mathcal{O}(\log^{1+\epsilon} m)$ time.

Proof. We use the following definitions. Let $m = |\mathcal{P}|$. Let $X_{\mathcal{P}}[1 \dots m_x]$ (resp. $Y_{\mathcal{P}}[1 \dots m_y]$) be a sorted array containing all different first (resp. second) coordinates of points in \mathcal{P} , i.e., such that $\{X_{\mathcal{P}}[i]\}_{i \in [1 \dots m_x]} = \{x : (x, y, w, \ell) \in \mathcal{P}\}$ (resp. $\{Y_{\mathcal{P}}[i]\}_{i \in [1 \dots m_y]} = \{y : (x, y, w, \ell) \in \mathcal{P}\}$). For any array of integers $A[1 \dots k]$ and any $a \in \mathbb{Z}$, we define $\mathrm{rank}(A, a) = |\{i \in [1 \dots k] : A[i] < a\}|$. Let \mathcal{P}_{ℓ} be the set \mathcal{P} with coordinates reduced to rank space and weights removed, i.e., $\mathcal{P}_{\ell} = \{(\mathrm{rank}(X_{\mathcal{P}}, x) + 1, \mathrm{rank}(Y_{\mathcal{P}}, y) + 1, \ell) : (x, y, w, \ell) \in \mathcal{P}\}$. Let \mathcal{P}_w be the set \mathcal{P} with coordinates reduced to rank space and labels removed, i.e., $\mathcal{P}_w = \{(\mathrm{rank}(X_{\mathcal{P}}, x) + 1, \mathrm{rank}(Y_{\mathcal{P}}, y) + 1, w) : (x, y, w, \ell) \in \mathcal{P}\}$. Note that by the assumption on \mathcal{P} , we have $|\mathcal{P}_{\ell}| = |\mathcal{P}_w| = m$. Let $P_{\mathrm{sort}}[1 \dots m]$ be an array containing all elements of \mathcal{P}_{ℓ} sorted by the second coordinate and, in case of ties, by the first coordinate.

Components The structure consists of five components:

- 1. The array $P_{\text{sort}}[1..m]$ in plain form using $\mathcal{O}(m)$ space.
- 2. The array $X_{\mathcal{P}}[1..m_x]$ in plain form using $\mathcal{O}(m_x) = \mathcal{O}(m)$ space.
- 3. The array $Y_{\mathcal{P}}[1..m_y]$ in plain form using $\mathcal{O}(m_y) = \mathcal{O}(m)$ space.
- 4. The data structure for weighted range counting from [Cha88] (where these queries are referred to as semigroup range searching) for the set \mathcal{P}_w . We use the variant using $\mathcal{O}(m)$ space that answers queries in $\mathcal{O}(\log^{2+\epsilon} m)$ time [Cha88, Table 1]. Note that as required by [Cha88], the coordinates of \mathcal{P}_w are in the rank space (i.e., $[1..|\mathcal{P}_w|]$) and there are no two points equal on both coordinates.
- 5. The data structure for range minimum queries from [Cha88] for the set \mathcal{P}_{ℓ} . We use the variant using $\mathcal{O}(m)$ space that answers queries in $\mathcal{O}(\log^{1+\epsilon} m)$ time [Cha88, Table 1]. Note that here again, as required by [Cha88], the coordinates of \mathcal{P}_{ℓ} are in the rank space, and there are no two points equal on both coordinates.

Implementation of queries The weighted range counting queries are answered as follows. Let $x_l, x_u \in \mathbb{Z}_{\geq 0}$ and $y_u \in \mathbb{Z}_{\geq 0}$. To compute weight-count_P (x_l, x_u, y_u) , we first reduce the query coordinates to rank space by performing binary search over X_P and Y_P to compute $x_l' = \operatorname{rank}(X_P, x_l) + 1$, $x_u' = \operatorname{rank}(Y_P, x_u) + 1$, and $y_u' = \operatorname{rank}(Y_P, y_u) + 1$. Observe, that then it holds weight-count_P $(x_l, x_u, y_u) = \sum_{(x,y,w) \in \mathcal{R}} w$, where $\mathcal{R} = \{(x,y,w) \in \mathcal{P}_w : x_l' \leq x < x_u' \text{ and } y < y_u'\}$. Thus, using the structure from [Cha88] on \mathcal{P}_w , the computation of weight-count_P (x_l, x_u, y_u) is performed analogously, except we first increment y_u . Lastly, the computation of weight-count_P (x_l, x_u, y_u) is reduced to weight-count_P (x_l, x_u, y_u) , where $y_u = Y_P[m_y] + 1$, i.e., we can immediately set $y_u' = m_y + 1$.

To answer range selection queries with arguments $x_l, x_u \in \mathbb{Z}_{\geq 0}$, and $r \in [1..\text{weight-count}_{\mathcal{P}}(x_l, x_u)]$, we first analogously reduce x_l and x_u in $\mathcal{O}(\log m)$ time to rank space coordinates x'_l and x'_u . We then perform a binary

search in $P_{\text{sort}}[1 ...m]$ with weighted range counting queries on \mathcal{P}_w as an oracle to find any $(x, y_u, \ell) \in \mathcal{P}_\ell$ such that $r \in (\sum_{(x,y,w)\in\mathcal{R}_1} w...\sum_{(x,y,w)\in\mathcal{R}_2} w]$, where $\mathcal{R}_1 = \{(x,y,w)\in\mathcal{P}_w: x_l' \leq x < x_u' \text{ and } y < y_u\}$ and $\mathcal{R}_2 = \{(x,y,w)\in\mathcal{P}_w: x_l' \leq x < x_u' \text{ and } y \leq y_u\}$. We then continue the binary search to locate any $(x',y',\ell')\in\mathcal{P}_\ell$ such that $y'=y_u$ and $x'\in[x_l'...x_u')$, and return $\ell'\in \text{weight-select}_{\mathcal{P}}(x_l,x_u,r)$ as the answer. In total, the selection query takes $\mathcal{O}(\log^{3+\epsilon} m)$ time.

To answer range minimum queries with arguments $x_l, x_u \in \mathbb{Z}_{\geq 0}$ and $y_l, y_u \in \mathbb{Z}_{\geq 0}$, we first reduce all coordinates in $\mathcal{O}(\log m)$ time to rank space coordinates x'_l, x'_u, y'_l, y'_u . Note that then it holds $\operatorname{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_u) = \min_{(x,y,\ell)\in\mathcal{R}}\ell$, where $\mathcal{R}=\{(x,y,\ell)\in\mathcal{P}_\ell: x'_l\leq x< x'_u \text{ and } y'_l\leq y< y'_u\}$. Thus, using the structure from [Cha88] on \mathcal{P}_ℓ , the computation of $\operatorname{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_u)$ takes $\mathcal{O}(\log^{1+\epsilon} m)$ time. The computation of $\operatorname{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_u)$, where $y_l=Y_{\mathcal{P}}[1]$ and $y_u=Y_{\mathcal{P}}[m_y]+1$, i.e., we can immediately set $y'_l=1$ and $y'_u=m_y+1$.

Construction algorithm Given the set \mathcal{P} , we first construct the arrays $X_{\mathcal{P}}$ and $Y_{\mathcal{P}}$ by scanning \mathcal{P} , and then sorting the resulting arrays and removing the duplicates. In total, this takes $\mathcal{O}(m \log m)$ time. We then construct \mathcal{P}_w and \mathcal{P}_ℓ . To this end, we need to reduce the coordinates of every point in \mathcal{P} to rank space. Overall, the construction $\mathcal{O}(m \log m)$ time. Next, we construct the structure for weighted range counting queries on \mathcal{P}_w and range minimum queries on \mathcal{P}_ℓ in $\mathcal{O}(m \log m)$ time using the algorithm described in [Cha88]. Finally, the array P_{sort} is computed from \mathcal{P}_ℓ in $\mathcal{O}(m \log m)$ time.

Proposition 6.2. Let $\epsilon > 0$ be a fixed constant. Given a set \mathcal{P} of m points with coordinates in $\mathbb{Z}_{\geq 0}$, we can in $\mathcal{O}(m \log m)$ time construct a structure of size $\mathcal{O}(m)$ supporting the following queries on \mathcal{P} :

- Weighted range counting queries in $\mathcal{O}(\log^{2+\epsilon} m)$ time,
- Weighted range selection queries in $\mathcal{O}(\log^{3+\epsilon} m)$ time,
- Range minimum queries in $\mathcal{O}(\log^{1+\epsilon} m)$ time.

Proof. Observe, that if there exists $p_1 = (x_1, y_2, w_1, \ell_1)$ and $p_2 = (x_2, y_2, w_2, \ell_2)$ in \mathcal{P} such that $p_1 \neq p_2$ and $(x_1, y_1) = (x_2, y_2)$ then, letting $\mathcal{P}' = \mathcal{P} \setminus \{p_1, p_2\} \cup \{(x_1, y_1, w_1 + w_2, \min(\ell_1, \ell_2))\}$, the answers to range counting and range minimum or \mathcal{P} and \mathcal{P}' are the same, and for every $x_l, x_u \in \mathbb{Z}_{\geq 0}$ and $r \in [1 ... \text{weight-count}_{\mathcal{P}}(x_l, x_u)]$, weight-select $\mathcal{P}(x_l, x_r, r) \neq \emptyset$ holds if and only if weight-select $\mathcal{P}'(x_l, x_r, r) \neq \emptyset$, which implies that we can use \mathcal{P}' instead of \mathcal{P} to execute range selection queries, since they return an arbitrary element of weight-select $\mathcal{P}(x_l, x_r, r)$. Let thus $\mathcal{P}_{\text{unique}}$ denote a set obtained by repeatedly merging pairs points in \mathcal{P} that are equal on both coordinates, as described above, until there are no such pairs. Formally, $\mathcal{P}_{\text{unique}}$ is such that

- 1. For every $(x, y, w, \ell) \in \mathcal{P}$, there exists $(x', y', w', \ell') \in \mathcal{P}_{unique}$ with (x', y') = (x, y).
- 2. For every $(x, y, w, \ell) \in \mathcal{P}_{\text{unique}}$, it holds $w = \sum_{(x', y', w', \ell') \in \mathcal{R}} w'$ and $\ell = \min_{(x', y', w', \ell') \in \mathcal{R}} \ell'$, where $\mathcal{R} = \{(x', y', w', \ell') \in \mathcal{P} : (x, y) = (x', y')\}.$

The data structure consists of a single component, namely, the structure from Proposition 6.1 for $\mathcal{P}_{\text{unique}}$. Letting $m' = |\mathcal{P}_{\text{unique}}|$, we have $m' \leq m$, and hence it needs $\mathcal{O}(m') = \mathcal{O}(m)$ space.

Queries are answered using the structure from Proposition 6.1. The correctness follows by the above discussion, and the time complexities follow by $m' \le m$ and by Proposition 6.1.

To construct $\mathcal{P}_{\text{unique}}$, we first sort in $\mathcal{O}(m \log m)$ time the input set \mathcal{P} by the first coordinate and, in case of ties, by the second one. The set $\mathcal{P}_{\text{unique}}$ is then easily obtained in $\mathcal{O}(m)$ time. In $\mathcal{O}(m' \log m') = \mathcal{O}(m \log m)$ we then construct the structure from Proposition 6.1 for $\mathcal{P}_{\text{unique}}$.

6.2 String-String Coordinates

Definition 6.3. Let $T \in \Sigma^n$ and $P \subseteq [1 ... n]$. For every $q \ge 1$, we define

$$\operatorname{StrStrPoints}_q(P,T) := \{(\overline{T^{\infty}[i-q\mathinner{.\,.} i)}, T^{\infty}[i\mathinner{.\,.} i+q), c(i), m(i)) : i \in P\},$$

where c(i) and m(i) are defined as follows:

• $c(i) = |\{i' \in [1 ... n] : T^{\infty}[i' - q ... i' + q) = T^{\infty}[i - q ... i + q)\}|,$ • $m(i) = \min\{i' \in [1 ... n] : T^{\infty}[i' - q ... i' + q) = T^{\infty}[i - q ... i + q)\}.$

Proposition 6.4. Let $T \in \Sigma^n$, $c = \max \Sigma$, $q \ge 1$, and $P \subseteq [1..n]$ be a set of |P| = p positions in T. Let $\epsilon > 0$ be a fixed constant and assume that we can compare any two substrings of T^{∞} or $\overline{T^{\infty}}$ (specified with their starting positions and lengths) in $\mathcal{O}(t_{\text{cmp}})$ time. There exists a data structure of size $\mathcal{O}(p)$ that, denoting $\mathcal{P} = \text{StrStrPoints}_q(P,T)$ (Definition 6.3), provides support for the following queries:

- 1. Given any $i \in [1..n]$ and $q_l, q_r \geq 0$, return weight-count_{\mathcal{D}} (x_l, x_u, y_l) and weight-count_{\mathcal{D}} (x_l, x_u, y_u) in $\mathcal{O}(\log^{2+\epsilon} n + t_{\text{cmp}} \log n)$ time, where $\overline{x_l} = T^{\infty}[i q_l ... i)$, $x_u = x_l c^{\infty}$, $y_l = T^{\infty}[i ... i + q_r)$, $y_u = y_l c^{\infty}$.
- 2. Given any $i \in [1..n], q_l \geq 0$, and $r \in [1..\text{weight-count}_{\mathcal{P}}(x_l, x_u)]$ (where $\overline{x_l} = T^{\infty}[i q_l..i)$ and $x_u = x_l c^{\infty}$), return any $j \in \text{weight-select}_{\mathcal{P}}(x_l, x_u, r)$ in $\mathcal{O}(\log^{3+\epsilon} n + t_{\text{cmp}} \log n)$ time.
- 3. Given any $i \in [1..n]$ and $q_l, q_r \ge 0$, return $\operatorname{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_u)$ in $\mathcal{O}(\log^{1+\epsilon} n + t_{\operatorname{cmp}} \log n)$ time, where $\overline{x_l} = T^{\infty}[i q_l ..i)$, $x_u = x_l c^{\infty}$, $y_l = T^{\infty}[i ..i + q_r)$, $y_u = y_l c^{\infty}$.

Furthermore, assuming that for any substring S of T (specified as above), we can compute |Occ(S,T)| and min Occ(S,T) in $O(t_{count})$ and $O(t_{minocc})$ time (respectively), then given the values q, ϵ , and the set P, we can construct the above data structure in $O(p \cdot (t_{cmp} \log n + t_{count} + t_{minocc}))$ time.

Proof. We use the following definitions. For every $i \in [1 ... n]$, denote $L(i) = \overline{T^{\infty}[i-q..i)}$ and $R(i) = T^{\infty}[i..i+q)$. Let $P_L[1...p]$ (resp. $P_R[1...p]$) be an array containing all positions in P ordered according to the reversed length-q left context (resp. according to the length-q right context) in T^{∞} , i.e., such that $\{P_L[i]\}_{i\in[1...p]} = P$ (resp. $\{P_R[i]\}_{i\in[1...p]} = P$) and that for every $i,i' \in [1...p]$, i < i' implies $L(P_L[i]) \leq L(P_L[i'])$ (resp. $R(P_R[i]) \leq R(P_R[i'])$). Let also $S_L[1...p]$ (resp. $S_R[1...p]$) be an array of strings corresponding to P_L and P_R , i.e., defined by $S_L[i] = L(P_L[i])$ (resp. $S_R[i] = R(P_R[i])$). For any array A[1...k] of strings over alphabet Σ , and any string $Q \in \Sigma^*$, we define rank $(A,Q) = |\{i \in [1...k] : A[i] \prec Q\}|$. We then define \mathcal{P}_{int} as \mathcal{P} with coordinates reduced from strings to integers, i.e., $\mathcal{P}_{\text{int}} = \{(\text{rank}(S_L, x) + 1, \text{rank}(S_R, y) + 1, w, \ell) : (x, y, w, \ell) \in \mathcal{P}\}$.

Components The structure consists of three components:

- 1. The array $P_L[1..p]$ in plain form using $\mathcal{O}(p)$ space.
- 2. The array $P_R[1..p]$ in plain form using $\mathcal{O}(p)$ space.
- 3. The data structure from Proposition 6.2 for \mathcal{P}_{int} using $\mathcal{O}(|\mathcal{P}_{int}|) = \mathcal{O}(|\mathcal{P}|) = \mathcal{O}(p)$ space.

Implementation of queries

- 1. First, we compute $y_l' = \operatorname{rank}(S_R, y_l) + 1$ and $y_u' = \operatorname{rank}(S_R, y_u) + 1$. To do this, we perform binary search over array P_R with the help of the oracle from the claim to compare substrings of T^{∞} . This lets us obtain y_l' . The computation of y_u' is slightly different since $y_u = y_l c^{\infty}$ is not a substring of T^{∞} or T^{∞} . To compute y_u' , we utilize the fact that for every substring S of T^{∞} it holds $S \prec y_l c^{\infty}$ if and only if $S \prec y_l$ or y_l is a prefix of S. The first condition is easy to check using the oracle. The second is equivalent to checking that $|y_l| \leq |S|$ and $y_l = S[1..|y_l|]$, which is again the query answered by the oracle. After computing y_l' and y_u' , we analogously compute $x_l' = \operatorname{rank}(S_L, x_l) + 1$ and $x_u' = \operatorname{rank}(S_L, x_u) + 1$ by binary search over P_L . In total, this takes $\mathcal{O}(t_{\rm cmp} \log n)$ time. Observe now that it holds weight-count $_{\mathcal{P}}(x_l, x_u, y_l) = \operatorname{weight-count}_{\mathcal{P}_{\rm int}}(x_l', x_u', y_l')$ and weight-count $_{\mathcal{P}}(x_l, x_u, y_u) = \operatorname{weight-count}_{\mathcal{P}_{\rm int}}(x_l', x_u', y_u')$. Thus, using the structure from Proposition 6.2, the rest of the query takes $\mathcal{O}(\log^{2+\epsilon} n)$ time. In total, we spend $\mathcal{O}(\log^{2+\epsilon} n + t_{\rm cmp} \log n)$ time.
- 2. First, as explained above, we compute $x_l' = \operatorname{rank}(S_L, x_l) + 1$ and $x_u' = \operatorname{rank}(S_L, x_u) + 1$ in $\mathcal{O}(t_{\text{cmp}} \log n)$ time. Observe that then weight-select_{$\mathcal{P}(x_l, x_u, r) = \text{weight-select}_{\mathcal{P}_{\text{int}}}(x_l', x_u', r)$. Thus, using the structure from Proposition 6.2, the rest of the query takes $\mathcal{O}(\log^{3+\epsilon} n)$ time. In total, we spend $\mathcal{O}(\log^{3+\epsilon} n + t_{\text{cmp}} \log n)$ time.}
- 3. First, compute in $\mathcal{O}(t_{\text{cmp}}\log n)$ time the values $x_l' = \text{rank}(S_L, x_l) + 1$, $x_u' = \text{rank}(S_L, x_u) + 1$, $y_l' = \text{rank}(S_R, y_l) + 1$, and $y_u' = \text{rank}(S_R, y_u) + 1$. Then, it holds $\text{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_r) = \text{r-min}_{\mathcal{P}_{\text{int}}}(x_l', x_u', y_l', y_u')$.

Thus, using the structure from Proposition 6.2, the rest of the query takes $\mathcal{O}(\log^{1+\epsilon} n)$ time. In total, we spend $\mathcal{O}(\log^{1+\epsilon} n + t_{\text{cmp}} \log n)$ time.

Construction algorithm First, we compute the arrays $P_L[1..p]$ and $P_R[1..p]$. In each case, we first initialize the array to contain all positions in P, and then perform comparison based sorting using either L(j) or R(j) as the key of each position j and the oracle to compare any two keys. In total, this takes $\mathcal{O}(t_{\rm cmp} \cdot p \log p)$ time. Next, we construct the set $\mathcal{P}_{\rm int}$. For each position $i \in P$, we first construct a tuple $({\sf rank}(S_L,L(i))+1,{\sf rank}(S_R,R(i))+1,c(i),m(i))$, where c(i) and m(i) are as in Definition 6.3. The first two elements in the tuple are computed using binary search over P_L and P_R , and the oracle for substring comparisons in $\mathcal{O}(t_{\rm cmp}\log n)$. The third and fourth elements are computed as follows. First, we check if $T^{\infty}[i-q\ldots i+q)$ overlaps T[n] (which holds if and only if i-q<1 or i+q>n). If so, then c(i)=1 and m(i)=i. Otherwise, we have $c(i)=|\mathrm{Occ}(T[i-q\ldots i+q),T)|$ and $m(i)=\min\mathrm{Occ}(T[i-q\ldots i+q),T)$, which we compute in $\mathcal{O}(t_{\rm count}+t_{\rm minocc})$ time. Over all points, we spend $\mathcal{O}(p\cdot (t_{\rm cmp}\log T+t_{\rm count}+t_{\rm minocc}))$ time. We collect all the resulting tuples in the array, which we then sort lexicographically in $\mathcal{O}(p\log p)$ time. Finally, we scan the array one last time and eliminate duplicates. The resulting array contains $\mathcal{P}_{\rm int}$. In $\mathcal{O}(p\log p)$ time we then construct the structure from Proposition 6.2. In total, we spend $\mathcal{O}(p\cdot (t_{\rm cmp}\log n+t_{\rm count}+t_{\rm minocc}))$ time.

6.3 Integer-String Coordinates

Definition 6.5. Let $T \in \Sigma^n$ and $P \subseteq [1 ... n] \times \mathbb{Z}_{>0}$. For every $q \ge 1$, we define

$$IntStrPoints_q(P,T) := \{(h, T^{\infty}[i ... i + q), c(i), m(i)) : (i, h) \in P\},\$$

where c(i) and m(i) are defined as follows:

```
 \begin{array}{l} \bullet \ c(i) = |\{i' \in [1 \mathinner{.\,.} n] : T^{\infty}[i' - q \mathinner{.\,.} i' + q) = T^{\infty}[i - q \mathinner{.\,.} i + q)\}|, \\ \bullet \ m(i) = \min\{i' \in [1 \mathinner{.\,.} n] : T^{\infty}[i' - q \mathinner{.\,.} i' + q) = T^{\infty}[i - q \mathinner{.\,.} i + q)\}. \end{array}
```

Proposition 6.6. Let $T \in \Sigma^n$, $c = \max \Sigma$, $q \ge 1$, and $P \subseteq [1..n] \times \mathbb{Z}_{\ge 0}$ be a set of |P| = p pairs. Let $\epsilon > 0$ be a fixed constant and assume that we can compare any two substrings of T^{∞} or $\overline{T^{\infty}}$ (specified with their starting positions and lengths) in $\mathcal{O}(t_{\text{cmp}})$ time. There exists a data structure of size $\mathcal{O}(p)$ that, denoting $\mathcal{P} = \text{IntStrPoints}_q(P,T)$ (Definition 6.5), provides support for the following queries:

- 1. Given any $i \in [1 \dots n]$, $x_l \ge 0$, and $q_r \ge 0$, return weight-count_P (x_l, n, y_l) and weight-count_P (x_l, n, y_u) in $\mathcal{O}(\log^{2+\epsilon} n + t_{\text{cmp}} \log n)$ time, where $y_l = T^{\infty}[i \dots i + q_r)$ and $y_u = y_l c^{\infty}$. Given any $i \in [1 \dots n]$ and $x_l \ge 0$, return the value weight-count_P (x_l, n) in $\mathcal{O}(\log^{2+\epsilon} n)$ time.
- 2. Given any $x_l \geq 0$ and $r \in [1..\text{weight-count}_{\mathcal{P}}(x_l, n)]$, return an element $j \in \text{weight-select}_{\mathcal{P}}(x_l, n, r)$ in $\mathcal{O}(\log^{3+\epsilon} n)$ time.
- 3. Given any $i \in [1..n]$, $x_l \ge 0$, and $q_r \ge 0$, return $\operatorname{r-min}_{\mathcal{P}}(x_l, n, y_l, y_u)$ in $\mathcal{O}(\log^{1+\epsilon} n + t_{\operatorname{cmp}} \log n)$ time, where $y_l = T^{\infty}[i..i + q_r)$ and $y_u = y_l c^{\infty}$. Given any $x_l \ge 0$, return $\operatorname{r-min}_{\mathcal{P}}(x_l, n)$ in $\mathcal{O}(\log^{1+\epsilon} n)$ time.

Furthermore, assuming that for any substring S of T (specified as above), we can compute |Occ(S,T)| and min Occ(S,T) in $O(t_{count})$ and $O(t_{minocc})$ time (respectively), then given the values q, ϵ , and the set P, we can construct the above data structure in $O(p \cdot (t_{cmp} \log n + t_{count} + t_{minocc}))$ time.

Proof. The data structure and its construction is similar to Proposition 6.4 and hence here we only describe the key differences. We use the following definitions. Denote $P' = \{i \in [1 ... n] : (i, h) \in P\}$ and p' = |P'|. Note that $p' \leq p$ and $|\mathcal{P}| \leq p$, but the inequalities can be strict. Let $P_L[1 ... p']$ (resp. $P_R[1 ... p']$) be an array containing all positions in P' ordered according to the left (resp. right) length-q contexts in T^{∞} . Let also $S_L[1 ... p']$ (resp. $S_R[1 ... p']$) be an array of strings defined by $S_L[i] = L(P_L[i])$ (resp. $S_R[i] = R(P_R[i])$), where $L(\cdot)$ (resp. $R(\cdot)$) is as in the proof of Proposition 6.4. We define \mathcal{P}_{int} as \mathcal{P} with second coordinates reduced from strings to integers, i.e., $\mathcal{P}_{\text{int}} = \{(x, \text{rank}(S_R, y) + 1, w, \ell) : (x, y, w, \ell) \in \mathcal{P}\}$ (where $\text{rank}(\cdot, \cdot)$ is defined as in the proof of Proposition 6.4).

Components The structure consists of two components:

- 1. The array $P_R[1..p']$ in plain form using $\mathcal{O}(p') = \mathcal{O}(p)$ space.
- 2. The data structure from Proposition 6.2 for \mathcal{P}_{int} using $\mathcal{O}(|\mathcal{P}_{int}|) = \mathcal{O}(|\mathcal{P}|) = \mathcal{O}(p)$ space.

Implementation of queries

- 1. First, in $\mathcal{O}(t_{\text{cmp}} \log n)$ time we compute $y'_l = \text{rank}(S_R, y_l) + 1$ and $y'_u = \text{rank}(S_R, y_u) + 1$ as in the proof of Proposition 6.4. Then, it holds weight-count_{\mathcal{P}} $(x_l, n, y_l) = \text{weight-count}_{\mathcal{P}_{int}}(x_l, n, y_l')$ and weight-count_P (x_l, n, y_u) = weight-count_{Pint} (x_l, n, y_u') . Thus, using Proposition 6.2, the rest of the query takes $\mathcal{O}(\log^{2+\epsilon} n)$ time. In total, we spend $\mathcal{O}(\log^{2+\epsilon} n + t_{\rm cmp} \log n)$ time. To compute weight-count_P (x_l, n) given $x_l \ge 0$, we observe that weight-count_P $(x_l, n) = \text{weight-count}_{P_{int}}(x_l, n)$. Thus, using Proposition 6.2, the query takes $\mathcal{O}(\log^{2+\epsilon} n)$ time.
- 2. Note that weight-select_{\mathcal{P}} $(x_l, n, r) = \text{weight-select}_{\mathcal{P}_{\text{int}}}(x_l, n, r)$. Thus, using the structure from Proposition 6.2, the query takes $\mathcal{O}(\log^{3+\epsilon} n)$ time.
- 3. First, compute in $\mathcal{O}(t_{\text{cmp}}\log n)$ time the values $y_l' = \text{rank}(S_R, y_l) + 1$ and $y_u' = \text{rank}(S_R, y_u) + 1$. Then, it holds $\operatorname{r-min}_{\mathcal{P}}(x_l, n, y_l, y_r) = \operatorname{r-min}_{\mathcal{P}_{\operatorname{int}}}(x_l, n, y_l', y_u')$. Thus, using the structure from Proposition 6.2, the rest of the query takes $\mathcal{O}(\log^{1+\epsilon} n)$ time. In total, we spend $\mathcal{O}(\log^{1+\epsilon} n + t_{\operatorname{cmp}} \log n)$ time. To compute $\operatorname{\mathsf{r-min}}_{\mathcal{P}}(x_l,n)$ given $x_l \geq 0$, we observe that $\operatorname{\mathsf{r-min}}_{\mathcal{P}}(x_l,n) = \operatorname{\mathsf{r-min}}_{\mathcal{P}_{\mathrm{int}}}(x_l,n)$. Thus, using Proposition 6.2, the query takes $\mathcal{O}(\log^{1+\epsilon} n)$ time.

Construction algorithm We begin by constructing an array containing elements of P'. Given P, this is easily done in $\mathcal{O}(p \log p)$ time. Next, we compute the arrays $P_L[1..p']$ and $P_R[1..p']$ as in the proof of Proposition 6.4 in $\mathcal{O}(t_{\text{cmp}} \cdot p' \log p')$ time. We then construct the set \mathcal{P}_{int} . To this end, for each $(i, h) \in P$, we construct a tuple $(\operatorname{rank}(S_L, L(i))+1, \operatorname{rank}(S_R, R(i))+1, i, c(i), m(i)),$ where c(i) and m(i) are as in Definition 6.5. The computation is performed as in the proof of Proposition 6.4 and takes $\mathcal{O}(p \cdot (t_{\text{cmp}} \log n + t_{\text{count}} + t_{\text{minocc}}))$ time. We then collect all the resulting tuples in the array, which we sort lexicographically in $\mathcal{O}(p \log p)$ time and with another scan remove the duplicates. The resulting array contains the set \mathcal{P}_{int} . In $\mathcal{O}(p \log p)$ time we then construct the structure from Proposition 6.2. In total, we spend $\mathcal{O}(p \cdot (t_{\text{cmp}} \log n + t_{\text{count}} + t_{\text{minocc}}))$ time.

Weighted Modular Constraint Queries

Let $\mathcal{I} \subseteq \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \times \mathbb{Z}$ be a finite set, where each tuple $(e, w, \ell) \in \mathcal{I}$ represents an interval [0..e] with an associated nonnegative endpoint e, a positive weight w, and an integer label ℓ . We assume that no two intervals share the same label.

Weighted modular constraint counting: Let $h \in \mathbb{Z}_{>0}$, $r \in [0..h)$, and $k_1, k_2 \in \mathbb{Z}_{>0}$ be such that $k_1 \leq k_2$.

- $\begin{array}{l} \bullet \ \operatorname{mod-count}_{\mathcal{I},h}(r,k_1,k_2) := \sum_{(e,w,\ell) \in \mathcal{I}} w \cdot |\{j \in [0\mathinner{\ldotp\ldotp\ldotp} e] : j \bmod h = r \bmod k_1 < \lfloor\frac{j}{h}\rfloor \le k_2\}|, \\ \bullet \ \operatorname{mod-count}_{\mathcal{I},h}(r,k_2) := \sum_{(e,w,\ell) \in \mathcal{I}} w \cdot |\{j \in [0\mathinner{\ldotp\ldotp\ldotp} e] : j \bmod h = r \bmod \lfloor\frac{j}{h}\rfloor \le k_2\}|, \\ \bullet \ \operatorname{mod-count}_{\mathcal{I},h}(r) := \sum_{(e,w,\ell) \in \mathcal{I}} w \cdot |\{j \in [0\mathinner{\ldotp\ldotp\ldotp} e] : j \bmod h = r\}|. \end{array}$

Weighted modular constraint selection: Let $h \in \mathbb{Z}_{>0}$, $r \in [0..h)$, and $c \in [1..mod-count_{\mathcal{I},h}(r)]$. We define mod-select_{\mathcal{I},h}(r,c):=k, where $k\in\mathbb{Z}_{>0}$ is the unique nonnegative integer satisfying $c\in\mathbb{Z}_{>0}$ $(\mathsf{mod\text{-}count}_{\mathcal{I},h}(r,k-1) \dots \mathsf{mod\text{-}count}_{\mathcal{I},h}(r,k)].$

Definition 7.1. Let $T \in \Sigma^n$ and $P \subseteq [1 ... n] \times \mathbb{Z}_{>0}$. For every $q \ge 1$, we define

WeightedIntervals_a
$$(P,T) := \{(e,c(i),m(i)) : (i,e) \in P\},$$

where c(i) and m(i) are defined as follows:

•
$$c(i) = |\{i' \in [1 ... n] : T^{\infty}[i' - q ... i' + q) = T^{\infty}[i - q ... i + q)\}|,$$

```
• m(i) = \min\{i' \in [1..n] : T^{\infty}[i' - q..i' + q) = T^{\infty}[i - q..i + q)\}.
```

Lemma 7.2. Let $T \in \Sigma^n$, $q \ge 1$, and $P \subseteq [1 ... n] \times \mathbb{Z}_{>0}$ be such that labels in $\mathcal{I} = \text{WeightedIntervals}_q(P, T)$ are unique. For every $h \in \mathbb{Z}_{>0}$, $r \in [0..h)$, and $k_1, k_2, k_3 \in \mathbb{Z}_{>0}$ such that $k_1 \leq k_2 \leq k_3$, it holds:

- 1. $\operatorname{\mathsf{mod-count}}_{\mathcal{I},h}(r,k_1,k_3) = \operatorname{\mathsf{mod-count}}_{\mathcal{I},h}(r,k_3) \operatorname{\mathsf{mod-count}}_{\mathcal{I},h}(r,k_1),$
- 2. $\mathsf{mod\text{-}count}_{\mathcal{T},h}(r,k_1,k_3) = \mathsf{mod\text{-}count}_{\mathcal{T},h}(r,k_1,k_2) + \mathsf{mod\text{-}count}_{\mathcal{T},h}(r,k_2,k_3)$.

Proof. By Definition 7.1, it follows that

$$\begin{aligned} \operatorname{mod-count}_{\mathcal{I},h}(r,k_1,k_3) &= \sum_{(e,w,\ell)\in\mathcal{I}} w \cdot |\{j \in [0\mathinner{\ldotp\ldotp\ldotp} e] : j \bmod h = r \bmod k_1 < \lfloor \frac{j}{h} \rfloor \leq k_3\}| \\ &= \sum_{(e,w,\ell)\in\mathcal{I}} w \cdot (|\{j \in [0\mathinner{\ldotp\ldotp\ldotp} e] : j \bmod h = r \bmod \lfloor \frac{j}{h} \rfloor \leq k_3\}| - \\ &\qquad |\{j \in [0\mathinner{\ldotp\ldotp\ldotp} e] : j \bmod h = r \bmod \lfloor \frac{j}{h} \rfloor \leq k_1\}|) \\ &= \sum_{(e,w,\ell)\in\mathcal{I}} w \cdot |\{j \in [0\mathinner{\ldotp\ldotp\ldotp} e] : j \bmod h = r \bmod \lfloor \frac{j}{h} \rfloor \leq k_3\}| - \\ &\qquad \sum_{(e,w,\ell)\in\mathcal{I}} w \cdot |\{j \in [0\mathinner{\ldotp\ldotp\ldotp} e] : j \bmod h = r \bmod \lfloor \frac{j}{h} \rfloor \leq k_1\}| \\ &= \operatorname{mod-count}_{\mathcal{I},h}(r,k_3) - \operatorname{mod-count}_{\mathcal{I},h}(r,k_1). \end{aligned}$$

By Item 1, we thus have

$$\begin{split} \mathsf{mod\text{-}count}_{\mathcal{I},h}(r,k_1,k_3) &= (\mathsf{mod\text{-}count}_{\mathcal{I},h}(r,k_2) - \mathsf{mod\text{-}count}_{\mathcal{I},h}(r,k_1)) + \\ & \qquad \qquad (\mathsf{mod\text{-}count}_{\mathcal{I},h}(r,k_3) - \mathsf{mod\text{-}count}_{\mathcal{I},h}(r,k_2)) \\ &= \mathsf{mod\text{-}count}_{\mathcal{I},h}(r,k_1,k_2) + \mathsf{mod\text{-}count}_{\mathcal{I},h}(r,k_2,k_3). \end{split}$$

Proposition 7.3. Let $T \in \Sigma^n$, $q \ge 1$, $h \in \mathbb{Z}_{>0}$, and $P \subseteq [1 ... n] \times [0 ... n]$ be a set of |P| = p pairs such that the labels in $\mathcal{I} = \text{WeightedIntervals}_q(P,T)$ are unique. Let $\epsilon > 0$ be a fixed constant. There exists a data structure of size $\mathcal{O}(p)$ that provides support for the following queries:

- 1. Given any $r \in [0..h)$ and $k_1, k_2 \in \mathbb{Z}_{\geq 0}$ satisfying $k_1 \leq k_2$, return the values mod-count_{\mathcal{I},h} (r, k_1, k_2) , $\operatorname{\mathsf{mod-count}}_{\mathcal{I},h}(r,k_1), \ \operatorname{\mathsf{and}} \ \operatorname{\mathsf{mod-count}}_{\mathcal{I},h}(r) \ \operatorname{\mathsf{in}} \ \mathcal{O}(\log^{2+\epsilon} n) \ \operatorname{\mathsf{time}}.$
- 2. Given any $r \in [0..h)$ and $c \in [1..mod\text{-count}_{\mathcal{I},h}(r)]$, return the value mod-select_{\mathcal{I},h}(r,c) in $\mathcal{O}(\log^{3+\epsilon} n)$

Furthermore, assuming that we can compare any two substrings of T^{∞} of $\overline{T^{\infty}}$ (specified with their starting positions and lengths) in $\mathcal{O}(t_{cmp})$ time and that for any substring S of T^{∞} (specified as above), we can compute $|\operatorname{Occ}(S,T)|$ and $\operatorname{min}\operatorname{Occ}(S,T)$ in $\mathcal{O}(t_{\operatorname{count}})$ and $\mathcal{O}(t_{\operatorname{minocc}})$ time (respectively), then given the values q, h, and ϵ , and the set P, we can construct the above data structure in $\mathcal{O}(p \cdot (t_{\rm cmp} \log n + t_{\rm count} + t_{\rm minocc}))$

Proof. Let $\mathcal{I} = \{(e_1, w_1, \ell_1), \dots, (e_{p'}, w_{p'}, \ell_{p'})\}$, where $p' = |\mathcal{I}|$ and $e_1 \leq e_2 \leq \dots \leq e_{p'}$. We also define $(e_0, w_0) = (0, 0)$. Note that $p' \leq p$. Note also that by Definition 7.1, for every $(i, e), (i', e') \in P$, i = i' implies that labels of the corresponding points in \mathcal{I} are equal. Thus, the assumption about unique labels implies $p \leq n$, and hence $p' \leq n$. For any $i \in [0 \dots p']$, denote $x_i = e_i \mod h$ and $y_i = \lfloor \frac{e_i}{h} \rfloor$. Let $Y[0 \dots p']$, $S_{\text{pref}}[0 \dots p']$, and $S_{\text{suf}}[0..p']$ be integer arrays defined such that for $i \in [0..p']$,

```
 \begin{split} \bullet & \ Y[i] = y_i, \\ \bullet & \ S_{\mathrm{pref}}[i] = \sum_{j=0}^i w_j \cdot y_j, \\ \bullet & \ S_{\mathrm{suf}}[i] = \sum_{j=i+1}^{p} w_j. \end{split}
```

Finally, let $\mathcal{P} = \{(x_i, y_i, w_i, \ell_i) : i \in [1..p']\}.$

Components The structure consists of two components:

- 1. The arrays Y[0..p'], $S_{\text{pref}}[0..p']$, and $S_{\text{suf}}[0..p']$ in plain form using $\mathcal{O}(p') = \mathcal{O}(p)$ space.
- 2. The data structure from Proposition 6.2 for the set \mathcal{P} . It needs $\mathcal{O}(p') = \mathcal{O}(p)$ space.

Implementation of queries

1. To compute $\mathsf{mod\text{-}count}_{\mathcal{I},h}(r,k_1)$, we proceed as follows. First, using binary search over the array Y, in $\mathcal{O}(\log p') = \mathcal{O}(\log n)$ time we compute $j = \max\{i \in [0..p'] : Y[i] \leq k_1\}$. Letting $\mathcal{J} = \{i \in [1..p'] : r \leq x_i < h \text{ and } 0 \leq y_i < k_1 + 1\}$, we then have:

$$\begin{aligned} \operatorname{\mathsf{mod-count}}_{\mathcal{I},h}(r,k_1) &= \sum_{i=0}^{p'} w_i \cdot \min(k_1,y_i) + \sum_{i \in \mathcal{J}} w_i \\ &= \sum_{i=0}^{j} w_i \cdot y_i + \sum_{i=j+1}^{p'} w_i \cdot k_1 + \sum_{i \in \mathcal{J}} w_i \\ &= S_{\operatorname{pref}}[j] + k_1 \cdot S_{\operatorname{suf}}[j] + \operatorname{\mathsf{weight-count}}_{\mathcal{P}}(r,h,0,k_1+1). \end{aligned}$$

The last term is computed in $\mathcal{O}(\log^{2+\epsilon} p') = \mathcal{O}(\log^{2+\epsilon} n)$ time using Proposition 6.2. In total, we thus spend $\mathcal{O}(\log^{2+\epsilon} n)$ time. The computation of mod-count_{\mathcal{I},h} (r,k_1,k_2) is reduced to mod-count_{\mathcal{I},h} (r,k_2) – mod-count_{\mathcal{I},h} (r,k_1) (Lemma 7.2(1)), and thus also takes $\mathcal{O}(\log^{2+\epsilon} n)$ time. To compute mod-count_{\mathcal{I},h}(r), we proceed as above, except we immediately set j=p'.

2. The value $\mathsf{mod\text{-}select}_{\mathcal{I},h}(r,c)$ is computed using binary search (in the range $[0 \dots n]$) and modular rank queries (implemented as above). Thus, the query takes $\mathcal{O}(\log^{3+\epsilon} n)$ time.

Construction algorithm First, for each pair $(i, e) \in P$, we construct a tuple (e, c(i), m(i)), where c(i) and m(i) are as in Definition 7.1. The computation is performed as in the proof of Proposition 6.4 and takes $\mathcal{O}(p \cdot (t_{\text{cmp}} \log n + t_{\text{count}} + t_{\text{minocc}}))$ time. We then collect all the resulting tuples in the array, which we sort lexicographically in $\mathcal{O}(p \log p)$ time and with another scan remove the duplicates. The resulting array contains the set \mathcal{I} (sorted by the first coordinate) and is of size p'. Using this array we compute the array S in $\mathcal{O}(p')$ time. Lastly, in $\mathcal{O}(p')$ time we construct the set \mathcal{P} , and then in $\mathcal{O}(p' \log p') = \mathcal{O}(p \log p)$ time we construct the structure from Proposition 6.2 for \mathcal{P} . In total, we spend $\mathcal{O}(p \cdot (t_{\text{cmp}} \log n + t_{\text{count}} + t_{\text{minocc}}))$ time. \square

8 Optimal Compressed Space SA and ISA Queries

For the duration of this section, we fix some $T \in \Sigma^n$ (where $\Sigma = [0..\sigma)$), such that T[n] does not occur in T[1..n). Let $\epsilon \in (0,1)$ be any fixed constant. We describe a data structure of size $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ that given any $j \in [1..n]$ (resp. $i \in [1..n]$) returns the value ISA[j] (resp. SA[i]) in $\mathcal{O}(\log^{4+\epsilon} n)$ time. Moreover, we show that given an LZ77 parsing of T, we can construct our data structure in $\mathcal{O}(\delta(T)\log^7 n)$ time.

8.1 Preliminaries

Definition 8.1 (τ -periodic and τ -nonperiodic patterns). Let $P \in \Sigma^m$ and $\tau \geq 1$. We say that P is τ -periodic if it holds $m \geq 3\tau - 1$ and $\text{per}(P[1...3\tau - 1]) \leq \frac{1}{3}\tau$. Otherwise, it is called τ -nonperiodic.

Definition 8.2 (Inverted lexicographic order). By \leq_{inv} we denote the inverted order on elements of Σ , i.e., $a \leq_{\text{inv}} b$ if and only if $b \leq a$ for $a, b \in \Sigma$. We then extend \leq_{inv} to inverted lexicographic order analogously to how \leq is extended to lexicographic order in Section 2.

Remark 8.3. Note that \leq_{inv} in Definition 8.2 is not the same as \succeq . For example, assuming $\Sigma = \{a, b, c\}$ with $a \prec b \prec c$, it holds $ab \leq abc$ and $ab \leq_{inv} abc$.

We now present equivalent characterizations of RangeBeg_{ℓ}(P,T), Pos^{beg}_{ℓ}(P,T), and Pos^{end}_{ℓ}(P,T). Although less intuitive than the original Definitions 2.1 and 3.1, they are useful in many contexts.

Lemma 8.4. Let $P \in \Sigma^*$ be such that no suffix of T is a proper prefix of P. Then, the following holds for every integer $\ell \geq 0$.

$$\begin{aligned} & \text{RangeBeg}_{\ell}(P,T) = |\{j' \in [1 \dots n] : T[j' \dots n] \prec P \ \ and \ \text{lcp}(P,T[j' \dots n]) < \ell\}|, \\ & \text{Pos}^{\text{beg}}_{\ell}(P,T) = \{j' \in [1 \dots n] : T[j' \dots n] \prec P \ \ and \ \text{lcp}(P,T[j' \dots n]) \in [\ell \dots 2\ell)\}, \\ & \text{Pos}^{\text{end}}_{\ell}(P,T) = \{j' \in [1 \dots n] : T[j' \dots n] \prec_{\text{inv}} P \ \ and \ \text{lcp}(P,T[j' \dots n]) \in [\ell \dots 2\ell)\}. \end{aligned}$$

Proof. Recall that RangeBeg_ℓ(P,T) = $|\{j' \in [1 \dots n] : T[j' \dots n] \prec P \text{ and } j' \notin \text{Occ}_ℓ(P,T)\}|$ and consider a position j' contributing to the right-hand side. By definition of $\text{Occ}_ℓ(P,T)$, we have $\text{lcp}(P,T[j' \dots n]) < \min(|P|,\ell) < \ell$. Consequently, RangeBeg_ℓ(P,T) ≤ $|\{j' \in [1 \dots n] : T[j' \dots n] \prec P \text{ and } \text{lcp}(P,T[j' \dots n]) < \ell\}|$. For a proof of the converse inequality, consider $j' \in [1 \dots n]$ such that $T[j' \dots n] \prec P$ and $\text{lcp}(P,T[j' \dots n]) < \ell$. The former condition implies that P is not a prefix of $T[j' \dots n]$, and thus $\text{lcp}(P,T[j' \dots n]) < |P|$. Together with the latter condition, this yields $\text{lcp}(P,T[j' \dots n]) < \min(|P|,\ell)$ and, in turn, $j' \notin \text{Occ}_ℓ(P,T)$. Consequently, $|\{j' \in [1 \dots n] : T[j' \dots n] \prec P \text{ and } \text{lcp}(P,T[j' \dots n]) < \ell\}| \leq \text{RangeBeg}_ℓ(P,T)$.

 $|\{j' \in [1 \dots n] : T[j' \dots n] \prec P \text{ and } \operatorname{lcp}(P,T[j' \dots n]) < \ell\}| \leq \operatorname{RangeBeg}_{\ell}(P,T).$ $\operatorname{Recall that } \operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T) = \{j' \in \operatorname{Occ}_{\ell}(P,T) : T[j' \dots n] \prec P \text{ and } j' \notin \operatorname{Occ}_{2\ell}(P,T)\} \text{ and consider a position } j' \in \operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T).$ $\operatorname{Due to } j' \in \operatorname{Occ}_{\ell}(P,T) \setminus \operatorname{Occ}_{2\ell}(P,T), \text{ we have } \min(|P|,\ell) \leq \operatorname{lcp}(P,T[j' \dots n]) < \min(|P|,2\ell).$ $\operatorname{In particular, } \min(|P|,\ell) < \min(|P|,2\ell) \text{ implies } |P| \geq \ell, \text{ so we derive } \ell = \min(|P|,\ell) \leq \operatorname{lcp}(P,T[j' \dots n]) < \min(|P|,2\ell).$ $\operatorname{In particular, } \min(|P|,\ell) \leq 2\ell. \text{ Consequently, } \operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T) \subseteq \{j' \in [1 \dots n] : T[j' \dots n] \prec P \text{ and } \operatorname{lcp}(P,T[j' \dots n]) \in [\ell \dots 2\ell)\}.$ For a proof of the converse inclusion, consider $j' \in [1 \dots n] \text{ such that } T[j' \dots n] \prec P \text{ and } \operatorname{lcp}(P,T[j' \dots n]) \in [\ell \dots 2\ell)\}.$ For a proof of the converse inclusion, consider $j' \in [1 \dots n] \text{ such that } T[j' \dots n] \prec P \text{ and } \operatorname{lcp}(P,T[j' \dots n]) \leq \ell \geq \min(|P|,\ell), \text{ so } j' \in \operatorname{Occ}_{\ell}(P,T).$ At the same time, $T[j' \dots n] \prec P \text{ implies that } P \text{ is not a prefix of } T[j' \dots n], \text{ so } \operatorname{lcp}(P,T[j' \dots n]) < |P|.$ Combined with $\operatorname{lcp}(P,T[j' \dots n]) < 2\ell, \text{ this implies } j' \notin \operatorname{Occ}_{2\ell}(P,T).$ Consequently, $\{j' \in [1 \dots n] : T[j' \dots n] \prec P \text{ and } \operatorname{lcp}(P,T[j' \dots n]) \in [\ell \dots 2\ell)\} \subseteq \operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T).$

Recall that $\operatorname{Pos}^{\operatorname{end}}_{\ell}(P,T) = \{j' \in \operatorname{Occ}_{\ell}(P,T) : T[j' ... n] \succ P \text{ and } j' \notin \operatorname{Occ}_{2\ell}(P,T)\}$ and consider a position $j' \in \operatorname{Pos}^{\operatorname{end}}_{\ell}(P,T)$. Due to $j' \in \operatorname{Occ}_{\ell}(P,T) \setminus \operatorname{Occ}_{2\ell}(P,T)$, we have $\min(|P|,\ell) \leq \operatorname{lcp}(P,T[j' ... n]) < \min(|P|,2\ell)$. In particular, $\min(|P|,\ell) < \min(|P|,2\ell)$ implies $|P| \geq \ell$, so we derive $\ell = \min(|P|,\ell) \leq \operatorname{lcp}(P,T[j' ... n]) < \min(|P|,2\ell) \leq 2\ell$. Moreover, $\operatorname{lcp}(P,T[j' ... n]) < |P|$ means that P is not a prefix of T[j' ... n], and thus $T[j' ... n] \succ P$ implies $T[j' ... n] \prec_{\operatorname{inv}} P$. Consequently, $\operatorname{Pos}^{\operatorname{end}}_{\ell}(P,T) \subseteq \{j' \in [1 ... n] : T[j' ... n] \prec_{\operatorname{inv}} P$ and $\operatorname{lcp}(P,T[j' ... n]) \in [\ell ... 2\ell)\}$. For a proof of the converse inclusion, consider $j' \in [1 ... n]$ such that $T[j' ... n] \prec_{\operatorname{inv}} P$ and $\operatorname{lcp}(P,T[j' ... n]) \in [\ell ... 2\ell)$. In particular, $\operatorname{lcp}(P,T[j' ... n]) \geq \ell \geq \min(|P|,\ell)$, so $j' \in \operatorname{Occ}_{\ell}(P,T)$. At the same time, $T[j' ... n] \prec_{\operatorname{inv}} P$ implies that P is not a prefix of T[j' ... n], so $\operatorname{lcp}(P,T[j' ... n]) < |P|$. Combined with $\operatorname{lcp}(P,T[j' ... n]) < 2\ell$, this implies $j' \notin \operatorname{Occ}_{2\ell}(P,T)$. Moreover, $T[j' ... n] \prec_{\operatorname{inv}} P$ implies $T[j' ... n] \succ P$ because T[j' ... n] is not a proper prefix of P (by the assumption in the lemma statement). Consequently, $\{j' \in [1 ... n] : T[j' ... n] \prec_{\operatorname{inv}} P$ and $\operatorname{lcp}(P,T[j' ... n]) \in [\ell ... 2\ell)\} \subseteq \operatorname{Pos}^{\operatorname{end}}_{\ell}(P,T)$.

Remark 8.5. We characterize $\operatorname{Pos}^{\operatorname{end}}_{\ell}(P,T)$ with the help of $\prec_{\operatorname{inv}}$, since due to the asymmetry of lexicographic order, $\{j' \in [1 \dots n] : T[j' \dots n] \succ P$ and $\operatorname{lcp}(P,T[j' \dots n]) \in [\ell \dots 2\ell)\}$ may contain elements of $\operatorname{Occ}(P,T)$.

 $\textbf{Lemma 8.6.} \ \ Let \ \ell \geq 1 \ \ and \ P \in \Sigma^+. \ \ Then, \ \mathrm{RangeBeg}_{2\ell}(P,T) = \mathrm{RangeBeg}_{\ell}(P,T) + \delta_{\ell}^{\mathrm{beg}}(P,T).$

 $\begin{array}{l} \textit{Proof.} \ \ \text{By combining Definition 3.1 and Lemma 8.4, we have } \ \text{RangeBeg}_{2\ell}(P,T) = |\{j' \in [1 \ldots n] : T[j' \ldots n] \prec P \ \text{and } \ \text{lcp}(P,T[j' \ldots n]) < 2\ell\}| = |\{j' \in [1 \ldots n] : T[j' \ldots n] \prec P \ \text{and } \ \text{lcp}(P,T[j' \ldots n]) < \ell\}| + |\{j' \in [1 \ldots n] : T[j' \ldots n] \prec P \ \text{and } \ \text{lcp}(P,T[j' \ldots n]) \in [\ell \ldots 2\ell)\}| = \text{RangeBeg}_{\ell}(P,T) + \delta_{\ell}^{\text{beg}}(P,T). \end{array}$

 $\textbf{Lemma 8.7.} \ \ Let \ \ell \geq 1 \ \ and \ j \in [1 \dots n]. \ \ Then, \ \mathrm{RangeBeg}_{2\ell}(j,T) = \mathrm{RangeBeg}_{\ell}(j,T) + \delta^{\mathrm{beg}}_{\ell}(j,T).$

Proof. Denote P = T[j ...n]. Note that, by definition, it holds $\operatorname{RangeBeg}_{\ell}(j,T) = \operatorname{RangeBeg}_{\ell}(P,T)$, $\operatorname{RangeBeg}_{2\ell}(j,T) = \operatorname{RangeBeg}_{2\ell}(P,T)$ (Definition 2.1), and $\delta_{\ell}^{\operatorname{beg}}(j,T) = \delta_{\ell}^{\operatorname{beg}}(P,T)$ (Definition 3.1). Thus, the claim follows by Lemma 8.6.

Lemma 8.8. Let $P \in \Sigma^*$ and $j \in [1..n]$. Denote m = |P|. Then:

- 1. For $0 \le \ell_1$, $P[1 ... \min(m, \ell_1)] \le T[j ... n]$ if and only if $T[j ... n] \succeq P$ or $lcp(P, T[j ... n]) \ge \ell_1$.
- 2. For $0 < \ell_2$, $T[j \dots n] \prec P[1 \dots \min(m, \ell_2)]$ if and only if $T[j \dots n] \prec P$ and $\operatorname{lcp}(P, T[j \dots n]) < \ell_2$.
- 3. For $0 \le \ell_1 < \ell_2$, $P[1 ... \min(m, \ell_1)] \le T[j ... n] < P[1 ... \min(m, \ell_2)]$ if and only if T[j ... n] < P and $lcp(P, T[j ... n]) \in [\ell_1 ... \ell_2)$.

Proof. 1. Denote $P_1 = P[1 .. \min(m, \ell_1)]$ and $\ell' = \operatorname{lcp}(P, T[j .. n])$. Let us first assume $P_1 \preceq T[j .. n]$. If $m \leq \ell_1$, then $|P_1| = \min(m, \ell_1) = m$, and hence $P_1 = P$. From the assumption we thus have $P = P_1 \preceq T[j .. n]$. Let us thus assume $m > \ell_1$. Then, $|P_1| = \ell_1$, and hence $P_1 = P[1 .. \ell_1]$. The assumption $P_1 \preceq T[j .. n]$ implies that either P_1 is a prefix of T[j .. n], or it holds $1 + \ell' \leq \ell_1$, $j + \ell' \leq n$, $P_1[1 .. \ell'] = T[j .. j + \ell')$, and $P_1[1 + \ell'] \prec T[j + \ell']$. Consider two cases.

- In the first case, we immediately obtain $\ell' = |P_1| = \ell_1$.
- In the second case, since P_1 is a prefix of P, we obtain $P[1..\ell'] = P_1[1..\ell'] = T[j..j + \ell')$ and $P[1+\ell'] = P_1[1+\ell'] \prec T[j+\ell']$. This implies $P \prec T[j..n]$.

Let us now assume that it holds $T[j ... n] \succeq P$ or $\ell' \geq \ell_1$. We consider both cases as follows:

- If $P \leq T[j ... n]$, then $P_1 \leq P \leq T[j ... n]$ follows, since P_1 is a prefix of P.
- Let us now assume $\ell' \geq \ell_1$. First, note that $m \geq \ell' \geq \ell_1$. Thus, $|P_1| = \min(m, \ell_1) = \ell_1$, i.e., $P_1 = P[1 ... \ell_1]$. On the other hand, the assumption $\ell' \geq \ell_1$ implies $P[1 ... \ell_1] = T[j ... j + \ell_1)$. We thus obtain $P_1 = P[1 ... \ell_1] = T[j ... j + \ell_1) \leq T[j ... n]$.
- 2. Denote $P_2 = P[1 ... \min(m, \ell_2)]$ and $\ell' = \text{lcp}(P, T[j ... n])$. Let us first assume $T[j ... n] \prec P_2$. We prove the two claims as follows:
 - Since P_2 is a prefix of P, we immediately obtain $T[j ... n] \prec P_2 \leq P$.
 - To show $\ell' < \ell_2$, suppose that we have $\ell' \ge \ell_2$. First, note that letting $P_3 = P[1 ... \ell']$ we have $P_3 = T[j ... j + \ell') \le T[j ... n]$. On the other hand, by $|P_2| = \min(m, \ell_2) \le \ell_2 \le \ell'$, we then also have $P_2 \le P_3$. Combining the two we thus obtain $P_2 \le P_3 \le T[j ... n]$, a contradiction.

Let us now assume that $T[j ... n] \prec P$ and $\ell' < \ell_2$. The assumption $T[j ... n] \prec P$ implies that either T[j ... n] is a proper prefix of P, or $j + \ell' \leq n$, $1 + \ell' \leq m$, $T[j ... j + \ell') = P[1 ... \ell']$, and $T[j + \ell'] \prec P[1 + \ell']$. Consider two cases:

- First, assume that T[j ...n] is a proper prefix of P. This implies that $\ell' = n j + 1$ and $m > \ell'$. On the other hand, we assumed $\ell' < \ell_2$. Thus, $|P_2| = \min(m, \ell_2) \ge \ell' + 1$, and hence $T[j ...n] = P[1 ...\ell'] \prec P[1 ...\ell' + 1] \preceq P_2$.
- Let us now assume that $j + \ell' \leq n$, $1 + \ell' \leq m$, $T[j ... j + \ell') = P[1 ... \ell']$, and $T[j + \ell'] \prec P[1 + \ell']$. We first observe that combining $1 + \ell' \leq m$ with the assumption $\ell' < \ell_2$ yields $|P_2| = \min(m, \ell_2) \geq \ell' + 1$. In particular, $P[1 ... \ell' + 1] \leq P_2$. On the other hand, by the assumption, we have $T[j ... n] \prec P[1 ... \ell' + 1]$. Putting the two observations together we thus obtain $T[j ... n] \prec P_2$.
- 3. Denote $P_1 = P[1 ... \min(m, \ell_1)]$, $P_2 = P[1 ... \min(m, \ell_2)]$, and $\ell' = \operatorname{lcp}(P, T[j ... n])$. Let us first assume $P_1 \preceq T[j ... n] \prec P_2$. By Lemma 8.8(2), it then follows that $T[j ... n] \prec P$ and $\ell' < \ell_2$. By subsequently applying Lemma 8.8(1), we thus obtain that it must additionally hold $\ell' \geq \ell_1$. Putting everything together, we obtain the claim. Let us now assume that $T[j ... n] \prec P$ and $\ell' \in [\ell_1 ... \ell_2)$. The claim then follows immediately by Lemma 8.8(1) and Lemma 8.8(2).

Lemma 8.9. Let $P_1, P_2 \in \Sigma^*$ be such that no nonempty suffix of T is a proper prefix of P_1 or P_2 . Let $j \in [1..n]$. Then:

- 1. For every $\ell \geq |P_1|$, $P_1 \leq T[j ... n]$ if and only if $P_1 \leq T^{\infty}[j ... j + \ell)$.
- 2. For every $\ell \geq |P_2|$, $T[j ... n] \prec P_2$ if and only if $T^{\infty}[j ... j + \ell) \prec P_2$.
- 3. For every $\ell \geq \max(|P_1|, |P_2|)$, $P_1 \leq T[j \dots n] \prec P_2$ if and only if $P_1 \leq T^{\infty}[j \dots j + \ell) \prec P_2$.

Proof. 1. Assume $P_1 \leq T[j ... n]$. Then, either P_1 is a prefix of T[j ... n], or there exists $\ell' \geq 0$ such that $\ell' < |P_1|, j + \ell' \leq n, P_1[1 ... \ell'] = T[j ... j + \ell')$, and $P_1[1 + \ell'] \prec T[j + \ell']$. We consider each case separately:

- If P_1 is a prefix of T[j ... n], then $j + |P_1| \le n + 1$, and hence $P_1 = T[j ... j + |P_1|) = T^{\infty}[j ... j + |P_1|) \le T^{\infty}[j ... j + \ell)$.
- Let us now assume that there exists $\ell' \geq 0$ such that $\ell' < |P_1|$, $j + \ell' \leq n$, $P_1[1 ... \ell'] = T[j ... j + \ell')$, and $P_1[1 + \ell'] \prec T[j + \ell']$. By definition of the lexicographic order, this implies $P_1 \prec T[j ... j + \ell')$, and hence we obtain $P_1 \prec T[j ... j + \ell') = T^{\infty}[j ... j + \ell') \leq T^{\infty}[j ... j + \ell)$.

Let us now assume that for some $\ell \geq |P_1|$, it holds $P_1 \leq T^{\infty}[j ... j + \ell)$. Consider two cases:

- First, assume $j + \ell \le n + 1$. Then, we obtain $P_1 \le T^{\infty}[j ... j + \ell) = T[j ... j + \ell) \le T[j ... n]$.
- Let us now assume $j + \ell > n + 1$. Denote $\ell' = \text{lcp}(P_1, T^{\infty}[j ... j + \ell))$. Since we assumed that no nonempty suffix of T is a proper prefix of P_1 , it must hold $j + \ell' \leq n$. By definition of the lexicographic order, $P_1 \leq T^{\infty}[j ... j + \ell)$ then implies $P_1[1 + \ell'] \prec T[j + \ell']$. Consequently, $P_1 \prec T[j ... j + \ell'] \leq T[j ... n]$.

2. Assume $T[j ... n] \prec P_2$. Observe that T[j ... n] not a prefix of P_2 , since otherwise we either have $T[j ... n] = P_2$, or some nonempty suffix of T is a proper prefix of P_2 . Consequently, letting $\ell' = \text{lcp}(T[j ... n], P_2)$, it holds $j + \ell' \leq n$, $\ell' < |P_2|$, and $T[j + \ell'] \prec P_2[1 + \ell']$. Since $T[j ... j + \ell']$ is a prefix of $T^{\infty}[j ... j + \ell)$, it thus follows by definition of the lexicographic order that $T^{\infty}[j ... j + \ell) \prec P_2[1 ... 1 + \ell'] \preceq P_2$.

Let us now assume $T^{\infty}[j ... j + \ell) \prec P_2$. Denote $\ell' = \operatorname{lcp}(T^{\infty}[j ... j + \ell), P_2)$. Observe that it holds $\ell' < |P_2|$ and $j + \ell' \leq n$. The inequality $\ell' < |P_2|$ holds since otherwise we would have $T^{\infty}[j ... j + \ell) \succeq P_2$. The inequality $j + \ell' \leq n$ then holds since otherwise T[j ... n] would be a nonempty suffix of T that is a proper prefix of P_2 . Consequently, by $\ell' < |P_2| \leq \ell$, we must have $T[j + \ell'] \prec P_2[1 + \ell']$. By definition of the lexicographic order, this implies $T[j ... n] \prec P_2[1 ... 1 + \ell'] \preceq P_2$.

3. The proof follows by combining the above two equivalences.

Lemma 8.10. Let $P \in \Sigma^*$. Let $j, j' \in [1 ... n]$ and $k \ge 0$ be such that $T^{\infty}[j ... j + k) = T^{\infty}[j' ... j' + k)$. Then, $j \in \operatorname{Occ}_k(P, T)$ if and only if $j' \in \operatorname{Occ}_k(P, T)$.

Proof. Let $j \in \operatorname{Occ}_k(P,T)$. We will prove that $j' \in \operatorname{Occ}_k(P,T)$ (the proof of the opposite implication follows by symmetry). If j = j', then the claim follows immediately. Let us thus assume $j \neq j'$. By the uniqueness of T[n] in T, the assumption $T^{\infty}[j \dots j + k) = T^{\infty}[j' \dots j' + k)$ implies $\operatorname{lcp}(T[j \dots n], T[j' \dots n]) = k'$, where $k' \geq k$. The assumption $j \in \operatorname{Occ}_k(P,T)$ implies $\operatorname{lcp}(T[j \dots n], P) = k''$, where $k'' \geq \min(|P|, k)$. We therefore obtain $\operatorname{lcp}(T[j' \dots n], P) \geq \min(\operatorname{lcp}(T[j' \dots n], T[j \dots n]), \operatorname{lcp}(T[j \dots n], P)) = \min(k', k'') \geq \min(k, \min(|P|, k)) = \min(|P|, k)$. Thus, $j' \in \operatorname{Occ}_k(P,T)$.

Lemma 8.11. For every $j, j' \in [1..n]$ and every $k \ge 0$, the following conditions are equivalent:

- $T^{\infty}[j \dots j+k) = T^{\infty}[j' \dots j'+k),$
- $\operatorname{Occ}_k(j,T) = \operatorname{Occ}_k(j',T)$,
- $j \in Occ_k(j', T)$,
- j = j' or $LCE_T(j, j') \ge k$.

Proof. Denote P = T[j ...n], P' = T[j' ...n], m = |P|, and m' = |P'|. Recall (Definition 2.1) that $\operatorname{Occ}_k(j,T)$ is defined as $\operatorname{Occ}_k(j,T) = \operatorname{Occ}_k(P,T) = \{j'' \in [1...n] : \operatorname{lcp}(P,T[j'' ...n]) \ge \min(m,k)\}$. Analogously, $\operatorname{Occ}_k(j',T) = \operatorname{Occ}_k(P',T) = \{j'' \in [1..n] : \operatorname{lcp}(P',T[j'' ...n]) \ge \min(m',k)\}$. The four implications are proved as follows.

- Assume $T^{\infty}[j ...j + k) = T^{\infty}[j' ...j' + k)$. If j = j', then P = P', and we immediately obtain $\operatorname{Occ}_k(j,T) = \operatorname{Occ}_k(P,T) = \operatorname{Occ}_k(P',T) = \operatorname{Occ}_k(j',T)$. Let us thus assume $j \neq j'$. By the uniqueness of T[n] in T, we then have $\operatorname{LCE}_T(j,j') \geq k$. In other words, $\operatorname{lcp}(P,P') = k'$, where $k' \geq k$. In particular, $m \geq k$ and $m' \geq k$. Consider any $j'' \in \operatorname{Occ}_k(j,T)$. Then, $\operatorname{lcp}(P,T[j''..n]) \geq \min(m,k) = k$. We thus have $\operatorname{lcp}(P',T[j''..n]) \geq \min(\operatorname{lcp}(P',P),\operatorname{lcp}(P,T[j''..n])) = \min(k',k) = k = \min(m',k)$. We have thus proved $j'' \in \operatorname{Occ}_k(j',T)$. Hence, $\operatorname{Occ}_k(j,T) \subseteq \operatorname{Occ}_k(j',T)$. The inclusion $\operatorname{Occ}_k(j',T) \subseteq \operatorname{Occ}_k(j,T)$ follows by symmetry. Thus, $\operatorname{Occ}_k(j,T) = \operatorname{Occ}_k(j',T)$.
- Assume $\operatorname{Occ}_k(j,T) = \operatorname{Occ}_k(j',T)$. By definition of P, it holds $\operatorname{lcp}(T[j\mathinner{\ldotp\ldotp} n],P) = m \geq \min(m,k)$. Thus, $j\in\operatorname{Occ}_k(j,T)$. By the assumption, this implies $j\in\operatorname{Occ}_k(j',T)$.
- Assume $j \in \operatorname{Occ}_k(j',T)$. If j=j', the claim follows immediately. Let us thus assume $j \neq j'$. By $j \in \operatorname{Occ}_k(j',T)$, we then have $\operatorname{lcp}(P',T[j\mathinner{.\,.}n]) \geq \min(m',k)$. Equivalently, it holds $\operatorname{LCE}_T(j',j) \geq \min(m',k)$. Suppose that $m' \leq k$. Then, $\min(m',k) = m'$, and the assumption $\operatorname{LCE}_T(j',j) \geq m'$ implies that $T[j\mathinner{.\,.} j+m'-1] = T[j'\mathinner{.\,.} j'+m'-1] = T[j'\mathinner{.\,.} n]$. In particular, T[n] = T[j+m'-1] = T[j+(n-j'+1)-1] = T[n-(j'-j)]. Since $j \neq j'$, this contradicts the uniqueness of T[n] in T. We thus have m' > k. Consequently, $\operatorname{LCE}_T(j',j) \geq \min(m',k) = k$. We have thus proved that it holds j = j' or $\operatorname{LCE}_T(j,j') \geq k$.
- Assume j = j' or $LCE_T(j, j') \ge k$. If j = j', then we immediately obtain $T^{\infty}[j ... j + k) = T^{\infty}[j' ... j' + k)$. Let us thus assume $LCE_T(j, j') \ge k$. This implies $T^{\infty}[j ... j + k) = T[j ... j + k) = T[j' ... j' + k) = T^{\infty}[j' ... j' + k)$. In both cases, we thus obtain $T^{\infty}[j ... j + k) = T^{\infty}[j' ... j' + k)$.

8.2 The Index Core

8.2.1 Preliminaries

Lemma 8.12. For every $c \ge 1$, it holds

$$\sum_{i=0}^{\infty} \frac{1}{2^i} \mathsf{d}_{c2^i}(T) = \mathcal{O}(c\delta(T) \log \frac{cn \log \sigma}{\delta(T) \log n}).$$

Proof. Let $\mu = \lfloor \log \frac{\log \delta(T)}{c \log \sigma} \rfloor$ and $\nu = \lceil \log \frac{n}{\delta(T)} \rceil$, so that $2^{\mu} \leq \frac{\log \delta(T)}{c \log \sigma}$ and $2^{\nu} \geq \frac{n}{\delta(T)}$.

• For $i \in [0..\mu]$, we observe that $\mathsf{d}_{c2^i}(T) \leq \sigma^{c2^i}$, and hence

$$\sum_{i=0}^{\mu} \tfrac{1}{2^i} \mathsf{d}_{c2^i}(T) \leq \sum_{i=0}^{\mu} \mathsf{d}_{c2^i}(T) \leq \sum_{i=0}^{\mu} \sigma^{c2^i} \leq 2\sigma^{c2^{\mu}} \leq 2\delta(T).$$

• For $i \in (\mu ... \nu]$, we observe that $\mathsf{d}_{c2^i}(T) \leq c \cdot 2^i \cdot \delta(T)$, and hence

$$\begin{split} \sum_{i=\mu+1}^{\nu} \frac{1}{2^i} \mathsf{d}_{c2^i}(T) &\leq \sum_{i=\mu+1}^{\nu} c\delta(T) \leq c\delta(T) (\nu - \mu) \\ &= \mathcal{O}(c\delta(T) \log \frac{cn \log \sigma}{\delta(T) \log \delta(T)}) = \mathcal{O}(c\delta(T) \log \frac{cn \log \sigma}{\delta(T) \log n}). \end{split}$$

• For $i \in (\nu ... \infty)$, we observe that $\mathsf{d}_{c2^i}(T) \leq n$, and hence

$$\sum_{i=\nu+1}^{\infty} \tfrac{1}{2^i} \mathsf{d}_{c2^i}(T) \leq \sum_{i=\nu+1}^{\infty} \tfrac{n}{2^i} \leq \tfrac{n}{2^{\nu+1}} \sum_{i=0}^{\infty} \left(\tfrac{1}{2} \right)^i \leq \delta(T).$$

Thus, in total, we obtain $\sum_{i=0}^{\infty} \frac{1}{2^i} \mathsf{d}_{c2^i}(T) = \mathcal{O}(c\delta(T) \log \frac{cn \log \sigma}{\delta(T) \log n})$.

Lemma 8.13. Let $\tau \in [1..\lfloor \frac{n}{2} \rfloor]$. Let $j, j', j'' \in [1..n]$ be such that $j, j'' \in \mathsf{R}(\tau, T), \ j' \notin \mathsf{R}(\tau, T), \ and <math>j < j' < j''$. Then, it holds $j'' - j \ge 2\tau$.

Proof. Suppose that $j''-j<2\tau$. Denote $p_1=\operatorname{per}(T[j\mathinner{\ldotp\ldotp} j+3\tau-1))$ and $p_2=\operatorname{per}(T[j''\mathinner{\ldotp\ldotp} j''+3\tau-1))$. By $j,j''\in\mathsf{R}(\tau,T)$, we have $p_1,p_2\leq\frac13\tau$. Let us first assume $p_1\neq p_2$. By $p_1,p_2\leq\frac13\tau$, each of the substrings $T[j\mathinner{\ldotp\ldotp} j+3\tau-1)$ and $T[j''\mathinner{\ldotp\ldotp} j''+3\tau-1)$ can be extended into a run [KK99, Mai89] (i.e., a maximal substring $T[x\mathinner{\ldotp\ldotp\ldotp} y)$ of T satisfying $\operatorname{per}(T[x\mathinner{\ldotp\ldotp\ldotp} y))\leq\frac12(y-x)$). Let $T[x_1\mathinner{\ldotp\ldotp\ldotp} y_1)$ (resp. $T[x_2\mathinner{\ldotp\ldotp\ldotp\ldotp y_2})$) be the run extending the substring $T[j\mathinner{\ldotp\ldotp\ldotp} j+3\tau-1)$ (resp. $T[j''\mathinner{\ldotp\ldotp\ldotp} j''+3\tau-1)$). Note that since $p_1\neq p_2$, these two runs are different. It is known that $T[x_1\mathinner{\ldotp\ldotp\ldotp y_1})$ and $T[x_2\mathinner{\ldotp\ldotp\ldotp y_2})$ cannot overlap by p_1+p_2 or more (e.g., [Koc18, Fact 2.2.4]). However, $p_1+p_2\leq\frac23\tau$, and $j''-j<2\tau$ implies that they overlap by at least τ symbols, a contradiction. We thus must have $p_1=p_2$. Note, however, that by definition of a period, this implies that $T[j\mathinner{\ldotp\ldotp\ldotp j''+3\tau-1})$ has period p_1 . In particular, $\operatorname{per}(T[j'\mathinner{\ldotp\ldotp\ldotp\ldotp j'+3\tau-1}))\leq\frac13\tau$, which contradicts $j'\not\in\mathsf{R}(\tau,T)$. We therefore must have $j''-j\geq 2\tau$.

Lemma 8.14. Let $\tau \in [1 ... | \frac{n}{2} |]$. For every $c \geq 1$, it holds

$$|\mathcal{I}(\text{comp}_{c\tau}(\mathsf{R}(\tau,T),T))| \leq \frac{5}{2\tau} \left(\mathsf{d}_{8c\tau}(T) + 8c\tau \right) \leq 40c\delta(T).$$

Proof. Denote $(s_k, t_k)_{k \in [1..m']} = \mathcal{I}(\text{comp}_{c\tau}(\mathsf{R}(\tau, T), T))$. Recall that $\text{comp}_{c\tau}(\mathsf{R}(\tau, T), T) = \mathsf{R}(\tau, T) \cap \mathsf{C}(c\tau, T)$ (Definition 4.10). Consider any maximal subinterval $[j ...j + \ell)$ of $\mathsf{C}(c\tau, T)$, i.e., such that $[j ...j + \ell) \subseteq \mathsf{C}(c\tau, T)$ and $\{j - 1, j + \ell\} \cap \mathsf{C}(c\tau, T) = \emptyset$. By definition, for every $k \in [1 ...m]$, $s_k \in (j ...j + \ell)$ implies $s_k - 1 \notin \mathsf{R}(\tau, T)$. On the other hand, we have by definition that $s_k \in \mathsf{R}(\tau, T)$. Therefore, letting $P = \{s_k : k \in [1 ...m]\}$, for every $j' \in (j ...j + \ell)$, $j' \in P$ implies $j' \in \mathsf{R}(\tau, T)$ and $j' - 1 \notin \mathsf{R}(\tau, T)$. By Lemma 8.13, we thus have that for

every $j', j'' \in (j ... j + \ell) \cap P$, it holds $|j'' - j'| \ge 2\tau$. Consequently, $|(j ... j + \ell) \cap P| \le \lceil \frac{\ell - 1}{2\tau} \rceil \le 1 + \lfloor \frac{\ell - 1}{2\tau} \rfloor$ and thus $|[j ... j + \ell) \cap P| \le 2 + \lfloor \frac{\ell}{2\tau} \rfloor$. Let us now denote $(x_k, y_k)_{k \in [1...m]} = \mathcal{I}(\mathsf{C}(c\tau, T))$. By Lemma 4.7 and the above discussion, it holds

$$\begin{split} |\mathcal{I}(\text{comp}_{c\tau}(\mathsf{R}(\tau,T),T))| &= \sum_{i=1}^{m} |[x_i \dots x_i + y_i) \cap P| \le \sum_{i=1}^{m} \left(2 + \left\lfloor \frac{y_i}{2\tau} \right\rfloor\right) \le 2m + \frac{1}{2\tau} \sum_{i=1}^{m} y_i \\ &\le 2 \left(\frac{\mathsf{d}_{8c\tau}(T)}{c\tau} + 8\right) + \frac{1}{2\tau} \left(\mathsf{d}_{8c\tau}(T) + 8c\tau\right) \\ &= \left(\frac{2}{c\tau} + \frac{1}{2\tau}\right) \cdot \left(\mathsf{d}_{8c\tau}(T) + 8c\tau\right) \\ &\le \frac{5}{2\tau} \left(\mathsf{d}_{8c\tau}(T) + 8c\tau\right). \end{split}$$

To show the second part of the claim, recall that for every $q \in \mathbb{Z}_{>0}$, we have $\mathsf{d}_q(T) + q \leq q \cdot \delta(T) + q \leq 2q \cdot \delta(T)$. Thus, $\frac{5}{2\tau} \left(\mathsf{d}_{8c\tau}(T) + 8c\tau \right) \leq \frac{5}{2\tau} \cdot 16c\tau \cdot \delta(T) \leq 40c\delta(T)$.

Lemma 8.15. Let $\tau \in [3 ... \lfloor \frac{n}{2} \rfloor]$. Let $i \in [1 ... n]$, and $b \leq 2\tau$ be such that $[i ... i + b) \subseteq [1 ... n - 3\tau + 2]$. Let $p = \operatorname{per}(T[i+b ... i+3\tau-1))$. If $p > \frac{1}{3}\tau$, then $[i ... i+b) \cap R(\tau,T) = \emptyset$. Otherwise, letting $x \in [i ... i+b]$ and $y \in [i+3\tau-1 ... i+b+3\tau-2]$ be such that $\operatorname{per}(T[x ... y)) = p$ and y - x is maximized, it holds

$$[i\mathinner{.\,.} i+b)\cap\mathsf{R}(\tau,T)=\begin{cases}\emptyset & if\ y-x<3\tau-1,\\ [x\mathinner{.\,.} y-3\tau+1] & otherwise.\end{cases}$$

Proof. We show the first implication by contraposition. Assume that $[i\mathinner{.\,.} i+b)\cap\mathsf{R}(\tau,T)\neq\emptyset$, i.e., there exists $i'\in[i\mathinner{.\,.} i+b)$ such that $i'\in\mathsf{R}(\tau,T)$. We then have $\mathsf{per}(T[i'\mathinner{.\,.} i'+3\tau-1))\leq\frac{1}{3}\tau$. By i'< i+b and $i+3\tau-1\leq i'+3\tau-1$, $T[i+b\mathinner{.\,.} i+3\tau-1)$ is a substring of $T[i'\mathinner{.\,.} i'+3\tau-1)$ and thus $p\leq\frac{1}{3}\tau$.

Let us thus assume $p \leq \frac{1}{3}\tau$. We first show, again by contraposition, that $y - x < 3\tau - 1$ implies $[i ... i + b) \cap R(\tau, T)$. Let i' be as above. Note that by $\operatorname{per}(T[i' ... i' + 3\tau - 1)) \leq \frac{1}{3}\tau$ we then have $x \leq i'$ and $y' \geq i' + 3\tau - 1$. Consequently, $y - x \geq 3\tau - 1$. This concludes the proof of the first case. Let us now assume $y - x \geq 3\tau - 1$. We aim to show that $[i ... i + b) \cap R(\tau, T) = [x ... y - 3\tau + 1]$.

- Let $i' \in [x ... y 3\tau + 1]$. By $x \ge i$, we have $i' \ge i$. On the other hand, by $y \le i + b + 3\tau 2$, we have $i' \le y 3\tau + 1 \le i + b 1$. Thus, $i' \in [i ... i + b)$. To show $i' \in \mathsf{R}(\tau, T)$, note that by the assumption $i' \le y 3\tau + 1$, or equivalently, $i' + 3\tau 1 \le y$, we have that $T[i' ... i' + 3\tau 1)$ is a substring of T[x ... y). Thus, $\mathsf{per}(T[i' ... i' + 3\tau 1)) \le \mathsf{per}(T[x ... y)) = p \le \frac{1}{3}\tau$, i.e., $i' \in \mathsf{R}(\tau, T)$.
- Let us now assume $i' \in [i ... i + b) \cap \mathbb{R}(\tau, T)$. Denote $p' = \operatorname{per}(T[i' ... i' + 3\tau 1))$. By $i' \in \mathbb{R}(\tau, T)$, we have $p' \leq \frac{1}{3}\tau$. As observed above, $i' \in [i ... i + b)$ implies that $T[i + b ... i + 3\tau 1)$ is a substring of $T[i' ... i' + 3\tau 1)$. This implies that $T[i' ... i' + 3\tau 1)$ in addition to p, also has period p'. Clearly, we have $p' \leq p$. We claim that p' = p. For the proof by contradiction, suppose p' < p. Note that $p + p' \leq 2\lfloor \frac{1}{3}\tau \rfloor \leq \tau 1 \leq 3\tau 1 b$ (where the last inequality follows by $b \leq 2\tau$). Thus, by the weak periodicity lemma [FW65], $T[i + b ... i + 3\tau 1)$ also has period $d = \gcd(p', p)$. By p' < p and $d \mid p'$, this implies d < p, contradicting $p = \operatorname{per}(T[i + b ... i + 3\tau 1))$. Thus, p' = p. This implies $x \leq i'$ and $y \geq i' + 3\tau 1$ and hence $i' \in [x ... y 3\tau + 1]$.

Lemma 8.16. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$. Let $t, t' \in \mathbb{Z}_{>0}$ be such that $t \leq t'$. Let $j \in [1 ... n]$ and $j' \in \operatorname{Occ}_{3\tau-1}(j,T)$ be such that $j' = \min \operatorname{Occ}_t(j',T)$. Then, $j \in \mathsf{R}(\tau,T)$ holds if and only if $j' \in \operatorname{comp}_{t'}(\mathsf{R}(\tau,T),T)$ (Definition 4.10).

Proof. Assume $j \in \mathsf{R}(\tau,T)$. This implies $\operatorname{per}(T[j\mathinner{\ldotp\ldotp\ldotp} j+3\tau-1)) \leq \frac{1}{3}\tau$. Thus, by the uniqueness of T[n] in T, the substring $T[j\mathinner{\ldotp\ldotp\ldotp} j+3\tau-1)$ does not contain the symbol T[n] and hence $j\leq n-3\tau+1$. By $j'\in\operatorname{Occ}_{3\tau-1}(j,T)$, we thus obtain $j'\leq n-3\tau-1$. Consequently, $T[j\mathinner{\ldotp\ldotp\ldotp} j+3\tau-1)=T[j'\mathinner{\ldotp\ldotp\ldotp} j'+3\tau-1)$. This implies $\operatorname{per}(T[j'\mathinner{\ldotp\ldotp\ldotp} j'+3\tau-1))=\operatorname{per}(T[j\mathinner{\ldotp\ldotp\ldotp} j+3\tau-1))\leq \frac{1}{3}\tau$, i.e., $j'\in\mathsf{R}(\tau,T)$. Recall now that $\operatorname{comp}_{t'}(\mathsf{R}(\tau,T),T)=\mathsf{R}(\tau,T)\cap\mathsf{C}(t',T)$ (see Construction 4.6 and Definition 4.10). By $t\leq t'$ and Lemma 4.4,

C(t',T) is a t-cover of T. Thus, $j'=\min Occ_t(j',T)$ implies that $[j'\ldots j'+t)\subseteq C(t',T)$ and hence $j'\in C(t',T)$. Combining with $j' \in \mathsf{R}(\tau,T)$, we thus obtain $j' \in \mathsf{R}(\tau,T) \cap \mathsf{C}(t',T) = \mathsf{comp}_{t'}(\mathsf{R}(\tau,T),T)$.

Let us now assume $j' \in \text{comp}_{t'}(\mathsf{R}(\tau,T),T)$. By definition, this implies $j' \in \mathsf{R}(\tau,T) \subseteq [1..n-3\tau+2]$. By $T^{\infty}[j'..j'+3\tau-1)=T^{\infty}[j..j+3\tau-1)$ (following from $j'\in \operatorname{Occ}_{3\tau-1}(j,T)$) and the uniqueness of T[n] in T, we have $j\in [1..n-3\tau+2]$. Thus, $T[j..j+3\tau-1)=T[j'..j'+3\tau-1)$ and hence $\operatorname{per}(T[j ... j + 3\tau - 1)) = \operatorname{per}(T[j' ... j' + 3\tau - 1)) \leq \frac{1}{3}\tau$. Consequently, $j \in R(\tau, T)$.

8.2.2 The Data Structure

Definitions For every $k \in [4..\lceil \log n \rceil)$, denote $\ell_k = 2^k$, $\tau_k = \lfloor \frac{\ell_k}{3} \rfloor$, and let $A_{\text{runs},k}[1..n_{\text{runs},k}]$ be an array containing the sequence $\mathcal{I}(\text{comp}_{14\tau_k}(\mathsf{R}(\tau_k,T),T))$ (Definitions 4.2 and 4.10) sorted lexicographically.

Components The index core, denoted CompSACore(T), consists of three components:

- 1. The structure from Theorem 5.24. It needs $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ space. 2. The structure from Theorem 5.25. It needs the same space as the first component.
- 3. For $k \in [4..\lceil \log n \rceil)$, we store the array $A_{\text{runs},k}[1..n_{\text{runs},k}]$ in plain form using $\mathcal{O}(n_{\text{runs},k})$ space. To bound the total space usage, first note that by Lemma 8.14, for every $k \in [4.. \lceil \log n \rceil)$, it holds $n_{\text{runs},k} \leq \frac{2}{5\tau_k} \left(\mathsf{d}_{112\tau_k}(T) + 112\tau_k \right)$. By utilizing that $3\tau_k \leq \ell_k \leq 4\tau_k$ and because $\mathsf{d}_q(T) + q$ is a non-decreasing function of $q \in \mathbb{Z}_{>0}$, we thus have $n_{\text{runs},k} \leq \frac{10}{\ell_k} \left(\mathsf{d}_{38\ell_k}(T) + 38\ell_k \right)$. Combining this with Lemma 8.12, we therefore obtain

$$\sum_{k=4}^{\lceil \log n \rceil - 1} n_{\text{runs},k} \le 10 \sum_{k=4}^{\infty} \frac{1}{\ell_k} \mathsf{d}_{38\ell_k}(T) + 380 \lceil \log n \rceil = \mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$$

where the in the last inequality we utilized that $\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n} = \Omega(\log n)$. The total space needed by all arrays is thus $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$.

In total, CompSACore(T) needs $\mathcal{O}(\delta(T)\log \frac{n\log \sigma}{\delta(T)\log n})$ space.

Basic Navigation Primitives

Proposition 8.17. Let $k \in [4 .. \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $j \in [1 ... n]$. Given CompSACore(T), the value k, the position j, and any $j' \in \operatorname{Occ}_{\ell}(j,T)$ satisfying $j' = \min \operatorname{Occ}_{2\ell}(j',T)$, we can in $\mathcal{O}(\log n)$ time determine if $j \in R(\tau, T)$.

Proof. First, note that by $3\tau - 1 \le \ell$, we have $\operatorname{Occ}_{\ell}(j,T) \subseteq \operatorname{Occ}_{3\tau-1}(j,T)$, and thus $j' \in \operatorname{Occ}_{3\tau-1}(j,T)$. On the other hand, we have $2\ell \le 14\tau$ (since for $\tau = \lfloor \frac{\ell}{3} \rfloor$ and $\ell \ge 16$, it holds $2\ell \le 7\tau$). Thus, by Lemma 8.16, $j \in \mathsf{R}(\tau,T)$ holds if and only if $j' \in \mathrm{comp}_{14\tau}(\mathsf{R}(\tau,T),T)$. We thus proceed as follows. If $n_{\mathrm{runs},k} = 0$, then $\operatorname{comp}_{14\tau}(\mathsf{R}(\tau,T),T)=\emptyset$ and hence we return that $j\not\in\mathsf{R}(\tau,T)$. Let us thus assume $n_{\mathrm{runs},k}>0$. If $A_{\text{runs},k}[1] > j'$, then by definition of $A_{\text{runs},k}$, we have $j' \notin \text{comp}_{14\tau}(\mathsf{R}(\tau,T),T)$ and thus again $j \notin \mathsf{R}(\tau,T)$. Otherwise, using binary search in $\mathcal{O}(\log n)$ time we compute the largest $i \in [1..n_{\text{runs},k}]$ such that letting $(p_i, t_i) = A_{\text{runs},k}[i]$, it holds $p_i \leq j'$. If $j' \in [p_i \dots p_i + t_i)$ then we have $j' \in \text{comp}_{14\tau}(\mathsf{R}(\tau,T),T)$ and hence $j \in \mathsf{R}(\tau, T)$. Otherwise we return $j \notin \mathsf{R}(\tau, T)$.

8.2.4 Construction Algorithm

Proposition 8.18. Given the LZ77 parsing of T, we can construct CompSACore(T) in $\mathcal{O}(\delta(T)\log^7 n)$ time.

Proof. We construct the components of CompSACore(T) (Section 8.2.2) as follows:

- 1. First, using Theorem 5.24, we construct the structure from Theorem 5.24 in $\mathcal{O}(\delta(T)\log^7 n)$ time.
- 2. Next, using Theorem 5.25, we construct the structure from Theorem 5.25 in $\mathcal{O}(\delta(T)\log^7 n)$ time.
- 3. We construct arrays $A_{\text{runs},i}$ as follows. First, using [KK20b, Theorem 6.7], in $\mathcal{O}(z(T)\log^2 n)$ time we construct a structure that, given any substring S of T (specified with its starting position and the length) in $\mathcal{O}(\log^3 n)$ time checks if $\operatorname{per}(S) \leq \frac{|S|}{2}$, and if so, returns $\operatorname{per}(S)$. Next, using [KK20b, Theorem 6.11],

in $\mathcal{O}(z(T)\log^4 n)$ time we construct a structure that, given any substring S of T (represented as above) in $\mathcal{O}(\log^3 n)$ time returns min $\operatorname{Occ}(S,T)$. For $i \in [4 ... \lceil \log n \rceil)$ we then execute the following algorithm:

- (a) Using Proposition 4.8 in $\mathcal{O}(z(T)\log^3 n)$ time we compute $(s_k, t_k)_{k \in [1..m]} = \mathcal{I}(\mathsf{C}(14\tau_i, T))$.
- (b) We then separately process each element of $(s_k, t_k)_{k \in [1..m]}$ in the order of increasing k (note that this corresponds to maximal blocks of $\mathsf{C}(14\tau_i, T)$ in the left-to-right order). For every $k \in [1..m]$, we split $[s_k \ldots s_k + t_k)$ into $\lceil \frac{t_k}{2\tau_i} \rceil$ subblocks of size $2\tau_i$ (except possibly the last one), and for each subblock $[j \ldots j + b)$, we determine using Lemma 8.15 if $[j \ldots j + b) \cap \mathsf{R}(\tau_i, T) \neq \emptyset$ and if so, compute positions $p, p' \in [j \ldots j + b]$ such that $[j \ldots j + b) \cap \mathsf{R}(\tau_i, T) = [p \ldots p')$. All subblocks within each block are processed left-to-right. Observe, that under this assumption, it is easy to generate subsequent elements of the array $A_{\text{runs},i}$ from the computed indexes p,p' for each subblock. It remains to explain how to efficiently implement the computation of p,p'.
 - i. First, in $\mathcal{O}(\log^3 n)$ time, we check if $\operatorname{per}(T[j+b\mathinner{\ldotp\ldotp} j+3\tau_i-1)) \leq \frac{3\tau_i-1-b}{2}$, and if so, we compute $\operatorname{per}(T[j+b\mathinner{\ldotp\ldotp\ldotp} j+3\tau_i-1))$. Observe that since $b\leq 2\tau$, it holds $3\tau_i-1-b\geq \tau_i-1$. Thus, $\operatorname{per}(T[j+b\mathinner{\ldotp\ldotp\ldotp} j+3\tau_i-1))>\frac{3\tau_i-1-b}{2}$ implies $\operatorname{per}(T[j+b\mathinner{\ldotp\ldotp\ldotp} j+3\tau_i-1))>\frac{\tau_i-1}{2}\geq \frac{1}{3}\tau_i$. By Lemma 8.15, we then have $[j\mathinner{\ldotp\ldotp\ldotp} j+b)\cap \mathsf{R}(\tau_i,T)=\emptyset$ and hence we can set p=p'=j. Let us thus assume that $p=\operatorname{per}(T[j+b\mathinner{\ldotp\ldotp\ldotp} j+3\tau_i-1))$ satisfies $p\leq \frac{3\tau_i-1-b}{2}$, and we obtained p.
 - thus assume that $p = \operatorname{per}(T[j+b\mathinner{.\,.} j+3\tau_i-1))$ satisfies $p \leq \frac{3\tau_i-1-b}{2}$, and we obtained p.

 ii. Next, we check if $p > \frac{1}{3}\tau_i$. If so, by Lemma 8.15 we again have $[j\mathinner{.\,.} j+b)\cap \mathsf{R}(\tau_i,T)=\emptyset$ and hence we return p=p'=j. Let us thus assume $p\leq \frac{1}{3}\tau_i$.
 - iii. Next, we determine the positions x and y by first computing in $\mathcal{O}(\log n)$ time the values $\delta_{\text{left}} = \text{LCE}_{\overline{T}}(n-j-b+2, n-j-b+2-p)$ and $\delta_{\text{right}} = \text{LCE}_{T}(j+3\tau_{i}-1, j+3\tau_{i}-1-p)$. We then have $x=j+b-\delta_{\text{left}}$ and $y=j+3\tau_{i}-1+\delta_{\text{right}}$. If $y-x<3\tau_{i}-1$, then by Lemma 8.15, we have $[j..j+b)\cap \mathsf{R}(\tau_{i},T)=\emptyset$, and thus we set p=p'=j. Otherwise, by Lemma 8.15, we set p=x and $p'=y-3\tau_{i}+2$.

Recall that block $[s_k \dots s_k + t_k)$ is split into $\lceil \frac{t_k}{2\tau_i} \rceil$ subblocks. The above procedure thus takes $\mathcal{O}((1 + \frac{t_k}{\tau_i}) \cdot \log^3 n)$ time. By Lemma 4.7, we have $m \leq 16\delta(T) = \mathcal{O}(\delta(T))$ and $\sum_{k \in [1...m]} t_k \leq 16 \cdot 14\tau_i \cdot \delta(T) = \mathcal{O}(\tau_i \delta(T))$. Over all $k \in [1...m]$ we thus spend $\mathcal{O}(\sum_{k \in [1...m]} (1 + \frac{t_k}{\tau_i}) \cdot \log^3 n) = \mathcal{O}(m \cdot \log^3 n + \frac{\tau_i \delta(T)}{\tau_i} \cdot \log^3 n) = \mathcal{O}(\delta(T) \log^3 n)$ time.

In total, the construction of all arrays $A_{\text{runs},i}$ takes $\mathcal{O}((z(T) + \delta(T))\log^3 n) = \mathcal{O}(\delta(T)\log^4 n)$ time. \square

8.3 The Nonperiodic Patterns and Positions

8.3.1 Preliminaries

Definition 8.19. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$. For every $Q \subseteq [1 ... n - 2\tau + 1]$ satisfying $Q \neq \emptyset$ and $\max Q \geq n - 3\tau + 2$, and every $j \in [1 ... n - 3\tau + 2]$, we denote $\operatorname{succ}_Q(j) = \min\{j' \in Q : j' \geq j'\}$. We then denote $\mathcal{D}(\tau, T, Q) := \{T[j ... \operatorname{succ}_Q(j) + 2\tau) : j \in [1 ... n - 3\tau + 2] \setminus \mathsf{R}(\tau, T)\}$.

Lemma 8.20. Let $\tau \geq 1$ and $P \in \Sigma^+$ be a τ -nonperiodic pattern. For every $P' \in \Sigma^+$, $lcp(P, P') \geq 3\tau - 1$ implies that P' is τ -nonperiodic.

Proof. By Definition 8.1, P being τ -nonperiodic implies that either $|P| < 3\tau - 1$, or $|P| \ge 3\tau - 1$ and $\operatorname{per}(P[1..3\tau - 1]) > \frac{1}{3}\tau$. From $\operatorname{lcp}(P, P') \ge 3\tau - 1$, we obtain $|P| \ge 3\tau - 1$. Thus, it holds $\operatorname{per}(P[1..3\tau - 1]) > \frac{1}{3}\tau$. Consequently, by $\operatorname{lcp}(P, P') \ge 3\tau - 1$, we obtain $\operatorname{per}(P'[1..3\tau - 1]) = \operatorname{per}(P[1..3\tau - 1]) > \frac{1}{3}\tau$, and hence P' is τ -nonperiodic.

Lemma 8.21. Let $\tau \geq 1$ be such that $3\tau - 1 \leq n$ and S be a τ -synchronizing set of T. Let also ℓ be such that $3\tau - 1 \leq \ell$ and $S_{\text{comp}} = \text{comp}_{\ell}(S, T)$ (Definition 4.10). Then,

- 1. It holds $S \neq \emptyset$ and $\max S \geq n 3\tau + 2$,
- 2. It holds $S_{comp} \neq \emptyset$ and $\max S_{comp} \geq n 3\tau + 2$.

Proof. 1. By the uniqueness of T[n] in T, it follows that $per(T[n-3\tau+2..n]) > \frac{1}{3}\tau$, i.e., $n-3\tau+2 \notin R(\tau,T)$. By the density property of S (Definition 2.4), we thus have $S \cap [n-3\tau+2..n-2\tau+2) \neq \emptyset$. This immediately implies both claims.

2. By the above, $S \neq \emptyset$. Let $s = \max S$. We will prove that $s \in S_{\text{comp}}$. Since $s \geq n - 3\tau + 2$, this will immediately imply both claims. Recall that by Definition 4.10, $S_{\text{comp}} = S \cap C(\ell, T)$. Thus, to show $s \in S_{\text{comp}}$, we need to prove $s \in C(\ell, T)$. By $s \ge n - 3\tau + 2$ and $\ell \ge 3\tau - 1$, it follows that $T^{\infty}[s ... s + \ell)$ contains the symbol T[n]. By the uniqueness of T[n] in T, it thus follows that the position s satisfies $s = \min \operatorname{Occ}_{\ell}(s, T)$. Since $C(\ell, T)$ is an ℓ -cover of T (Definition 4.10), it thus follows by Lemma 4.5 that $s \in C(\ell, T)$.

Lemma 8.22. Let $\tau \in [1 ... |\frac{n}{2}|]$, S be a τ -synchronizing set of T, and $j \in [1 ... n]$.

- 1. Let P be a τ -nonperiodic pattern satisfying $|P| \geq 3\tau 1$. Then, $j \in Occ_{3\tau-1}(P,T)$ implies that:
 - $j \in [1..n 3\tau + 2] \setminus R(\tau, T)$,
 - $T[j .. \operatorname{succ}_{S}(j) + 2\tau)$ is a prefix of P.
- 2. Let $j' \in [1 ... n 3\tau + 2] \setminus R(\tau, T)$. Then, $j \in Occ_{3\tau 1}(j', T)$ implies that:
 - $j \in [1..n 3\tau + 2] \setminus R(\tau, T)$,
 - $\operatorname{succ}_{S}(j) j = \operatorname{succ}_{S}(j') j'$.

Proof. 1. Denote m = |P|. By definition of $Occ_{3\tau-1}(P,T)$ and the assumption $m \geq 3\tau - 1$, it holds $lcp(T[j..n], P) \ge min(m, 3\tau - 1) = 3\tau - 1$. Consequently, we must have $j + 3\tau - 1 \le n + 1$, or equivalently, $j \in I$ $[1...n-3\tau+2]$. By Lemma 8.20, we then obtain that T[j...n] is τ -nonperiodic. Thus, $j \in [1...n-3\tau+2] \setminus R(\tau,T)$.

We now show the second claim. First, note that $succ_s(j)$ is well-defined by Lemma 8.21 (note that $3\tau - 1 \le n$ follows by $j \in [1 ... n - 3\tau + 2]$. To show the main claim, note that by the density property of S (Definition 2.4) and $j \in [1 ... n - 3\tau + 2] \setminus R(\tau, T)$, we have $S \cap [j ... j + \tau) \neq \emptyset$. Thus, $\operatorname{succ}_S(j) - j < \tau$, and thus letting $D = T[j .. \operatorname{succ}_{S}(j) + 2\tau)$, it holds $|D| = (\operatorname{succ}_{S}(j) - j) + 2\tau \le 3\tau - 1$. Consequently, by $lcp(T[j ... n], P) \geq 3\tau - 1$, we obtain that D is a prefix of P.

2. Denote P = T[j' ... n] and $m = |P'| \ge 3\tau - 1$. By $j' \notin R(\tau, T)$, P is τ -nonperiodic. Recall that, by definition, $Occ_{3\tau}(j',T) = Occ_{3\tau-1}(P,T)$. Thus, by $j \in Occ_{3\tau-1}(P,T)$ and Lemma 8.22(1), we obtain that $j \in [1 \dots n - 3\tau + 2] \setminus \mathsf{R}(\tau, T).$

We now show the second claim. The notation $succ_{S}(j)$ and $succ_{S}(j')$ is well-defined by an analogous argument as above. Denote $s = \text{succ}_{S}(j)$ and $s' = \text{succ}_{S}(j')$. By the above, we have $j, j' \in [1 ... n - 3\tau + 2] \setminus$ $R(\tau,T)$. Thus, by Definition 2.4, $S \cap [j ... j + \tau) \neq \emptyset$ and $S \cap [j' ... j' + \tau) \neq \emptyset$. On the other hand, note that $LCE_T(j, j') \ge 3\tau - 1$. By the consistency condition of S, we thus have that for every $\delta \in [0..\tau)$, $j + \delta \in S$ holds if and only if $j' + \delta \in S$. Applying this for $\delta \in [0 ... s - j]$, we thus obtain that $S \cap [j' ... j' + (s - j)] = \{j' + (s - j)\}$. Consequently, s' = j' + (s - j) which is equivalent to s - j = s' - j', i.e., the claim.

The Data Structure 8.3.2

Definitions For every $k \in [4 .. \lceil \log n \rceil)$, we denote $\ell_k = 2^k$ and $\tau_k = \lfloor \frac{\ell_k}{3} \rfloor$. Note that $\ell_k \in [16 .. n)$. By $\{\mathsf{S}_{\text{comp},k}\}_{k \in [4 .. \lceil \log n \rceil)}$, we denote the collection obtained using Proposition 5.30 for c = 14. For every $k \in [4..\lceil \log n \rceil)$, by S_k we denote the τ_k -synchronizing set of T satisfying $S_{\text{comp},k} = \text{comp}_{14\tau_k}(S_k,T)$ (Definition 4.10). Such S_k exists by Proposition 5.30. We also let $n_k = |S_{\text{comp},k}|$ and by $A_{\text{comp},k}[1...n_k]$ denote an array containing the elements of $S_{\text{comp},k}$ in sorted order.

Components The data structure, denoted CompSANonperiodic (T), to handle nonperiodic positions consists of three components:

- 1. The index core CompSACore(T) (Section 8.2.2). It needs $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ space. 2. For $k \in [4\mathinner{.\,.}\lceil\log n\rceil)$, we store the array $A_{\operatorname{comp},k}[1\mathinner{.\,.} n_k]$ in plain form using $\mathcal{O}(n_k)$ space. By Proposition
- tion 5.30, the total space is $\mathcal{O}(\sum_{k \in [4.. \lceil \log n \rceil)} n_k) = \mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$.

 3. For $k \in [4.. \lceil \log n \rceil)$, we store the structure from Proposition 6.4 for $P = \mathsf{S}_{\text{comp},k}$ and $q = 7\tau_k$. By Propositions 5.30 and 6.4, in total they need $\mathcal{O}(\sum_{k \in [4.. \lceil \log n \rceil)} n_k) = \mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$ space. In total, CompSANonperiodic(T) needs $\mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$ space.

8.3.3 Basic Combinatorial Properties

Lemma 8.23. Let $\tau \in [1..\lfloor \frac{n}{2} \rfloor]$. Let S be a τ -synchronizing set of T and let $S_{\text{comp}} = \text{comp}_{14\tau}(S, T)$ (Definition 4.10). Let $\mathcal{P} = \text{StrStrPoints}_{7\tau}(S_{\text{comp}}, T)$ (Definition 6.3). For any strings x_l, x_u, y_l, y_u , it holds:

- $1. \ \ \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l, x_u, y_l, y_u) = |\{s \in \mathsf{S} : x_l \underline{\preceq} \overline{T^{\infty}[s 7\tau..s)} \prec x_u \ \ and \ y_l \preceq T^{\infty}[s..s + 7\tau) \prec y_u\}|,$
- 2. weight-count $\frac{\preceq}{\mathcal{P}}(x_l, x_u, y_u) = |\{s \in \mathsf{S} : x_l \preceq \overline{T^{\infty}[s-7\tau..s)} \prec x_u \text{ and } T^{\infty}[s..s+7\tau) \preceq y_u\}|,$
- 3. weight-count_P $(x_l, x_u) = |\{s \in S : x_l \preceq \overline{T^{\infty}[s-7\tau..s)} \prec x_u\}|,$

Proof. 1. The proof consists of three steps labeled (a) through (c).

(a) Denote

$$Q = \{ T^{\infty}[s - 7\tau ... s + 7\tau) : s \in \mathsf{S}_{\text{comp}}, \ x_l \preceq \overline{T^{\infty}[s - 7\tau ... s)} \prec x_u \text{ and } y_l \preceq T^{\infty}[s ... s + 7\tau) \prec y_u \},$$

$$A = \{ i \in [1 ... n] : T^{\infty}[i - 7\tau ... i + 7\tau) \in Q \}.$$

In the first step, we prove that weight-count_p(x_l, x_u, y_l, y_u) = |A|. Let $\mathcal{R} = \{(x, y, w, \ell) \in \mathcal{P} : x_l \leq x < x_u \text{ and } y_l \leq y \leq y_u\}$. We begin by proving that for every $p = (x_p, y_p, w_p, \ell_p) \in \mathcal{R}$, it holds $T^{\infty}[\ell_p - 7\tau \dots \ell_p + 7\tau) \in Q$. By $p \in \mathcal{P}$ (see Definition 6.3), there exists $i_p \in S_{\text{comp}}$ such that $x = \overline{T^{\infty}[i_p - 7\tau \dots i_p)}$, $y = T^{\infty}[i_p \dots i_p + 7\tau)$, and ℓ_p satisfies $T^{\infty}[\ell_p - 7\tau \dots \ell_p + 7\tau) = T^{\infty}[i_p - 7\tau \dots i_p + 7\tau)$. By $p \in \mathcal{R}$, we thus have $x_l \leq x \leq x_u$, i.e., $x_l \leq \overline{T^{\infty}[i_p - 7\tau \dots i_p)} \leq x_u$ and $y_l \leq y \leq y_u$, i.e., $y_l \leq T^{\infty}[i_p \dots i_p + 7\tau) \leq y_u$. We have thus proved that $T^{\infty}[i_p - 7\tau \dots i_p + 7\tau) \in Q$. As observed above, however, it holds $T^{\infty}[\ell_p - 7\tau \dots \ell_p + 7\tau) = T^{\infty}[i_p - 7\tau \dots i_p + 7\tau)$. Thus, $T^{\infty}[\ell_p - 7\tau \dots \ell_p + 7\tau) \in Q$. Let $g: \mathcal{R} \to Q$ be a function defined so that for every $p = (x, y, w, \ell) \in \mathcal{R}$, it holds $g(p) = T^{\infty}[\ell - 7\tau \dots \ell + 7\tau)$. We prove that g is a bijection.

- Let $p_1, p_2 \in \mathcal{R}$ be such that $g(p_1) = g(p_2)$. We will show that $p_1 = p_2$. For $k \in \{1, 2\}$, denote $p_k = (x_k, y_k, w_k, \ell_k)$ and let $i_k \in \mathsf{S}_{\text{comp}}$ be such that $x_k = \overline{T^{\infty}[i_k 7\tau \dots i_k)}, y_k = T^{\infty}[i_k \dots i_k + 7\tau)$, and $T^{\infty}[\ell_k 7\tau \dots \ell_k + 7\tau) = T^{\infty}[i_k 7\tau \dots i_k + 7\tau)$. Let also $S_k = T^{\infty}[\ell_k 7\tau \dots \ell_k + 7\tau) = T^{\infty}[i_k 7\tau \dots i_k + 7\tau)$. Note that then $g(p_k) = S_k$. By the assumption, $S_1 = g(p_1) = g(p_2) = S_2$. Consequently, $\ell_1 = \min\{j \in [1 \dots n] : T^{\infty}[j 7\tau \dots j + 7\tau) = S_1\} = \min\{j \in [1 \dots n] : T^{\infty}[j 7\tau \dots j + 7\tau) = S_2\} = \ell_2$. Similarly, $w_1 = |\{j \in [1 \dots n] : T^{\infty}[j 7\tau \dots j + 7\tau) = S_1\}| = |\{j \in [1 \dots n] : T^{\infty}[j 7\tau \dots j + 7\tau) = S_2\}| = w_2$. We also have $x_1 = \overline{T^{\infty}[i_1 7\tau \dots i_1)} = \overline{S_1[1 \dots 7\tau]} = \overline{S_2[1 \dots 7\tau]} = \overline{T^{\infty}[i_2 7\tau \dots i_2)} = x_2$ and $y_1 = T^{\infty}[i_1 \dots i_1 + 7\tau) = S_1(7\tau \dots 14\tau] = S_2(7\tau \dots 14\tau) = T^{\infty}[i_2 \dots i_2 + 7\tau) = y_2$. We have thus proved $p_1 = p_2$.
- Let $X \in Q$. Then, there exists $s \in \mathsf{S}_{\text{comp}}$ such that $x_l \preceq \overline{T^{\infty}[s-7\tau..s)} \prec x_u, \ y_l \preceq T^{\infty}[s..s+7\tau) \prec y_u$, and $T^{\infty}[s-7\tau..s+7\tau) = X$. Note that by definition of \mathcal{P} , $s \in \mathsf{S}_{\text{comp}}$ implies that there exists $p = (x, y, w, \ell) \in \mathcal{P}$ such that $x = \overline{T^{\infty}[s-7\tau..s)}, \ y = T^{\infty}[s..s+7\tau)$, and ℓ satisfies $T^{\infty}[\ell-7\tau..\ell+7\tau) = T^{\infty}[s-7\tau..s+7\tau) = X$. By the above observations, we thus have $x_l \preceq x \prec x_u$ and $y_l \preceq y \prec y_u$. Thus, $p \in \mathcal{R}$. Since, as noted earlier, we have $T^{\infty}[\ell-7\tau..\ell+7\tau) = X$, we thus obtain $g(p) = T^{\infty}[\ell-7\tau..\ell+7\tau) = X$.

We have thus proved that g is a bijection. Observe now that, by definition (see Section 6), it holds weight-count_{\mathcal{P}}(x_l, x_u, y_l, y_u) = \sum_{(x,y,w,\ell) \in \mathcal{R}} w. Observe also (see Definition 6.3) that for every $(x, y, w, \ell) \in \mathcal{P}$, it holds $w = |\{j \in [1 ... n] : T^{\infty}[j - 7\tau ... j + 7\tau) = T^{\infty}[\ell - 7\tau ... \ell + 7\tau)\}|$. Thus, weight-count_{\mathcal{P}}(x_l, x_u, y_l, y_u) = \sum_{(x,y,w,\ell) \in \mathcal{R}} |\{j \in [1 ... n] : T^{\infty}[j - 7\tau ... j + 7\tau) = T^{\infty}[\ell - 7\tau ... \ell + 7\tau)\}|. Observe that since g is bijection, for any $(x_1, y_1, w_1, \ell_1), (x_2, y_2, w_2, \ell_2) \in \mathcal{R}$, $(x_1, y_1, w_1, \ell_1) \neq (x_2, y_2, w_2, \ell_2)$ implies $T^{\infty}[\ell_1 - 7\tau ... \ell_1 + 7\tau) \neq T^{\infty}[\ell_2 - 7\tau ... \ell_2 + 7\tau)$. Thus, in the above expression, no position $j \in [1 ... n]$ is accounted twice (i.e., for different elements of \mathcal{R}). On the other hand, g being a bijection also implies that for every $j \in [1 ... n]$, $T^{\infty}[j - 7\tau ... j + 7\tau) \in Q$ implies that there exists some $p = (x, y, w, \ell) \in \mathcal{R}$ such that $T^{\infty}[j - 7\tau ... j + 7\tau) = T^{\infty}[\ell - 7\tau ... \ell + 7\tau)$. Consequently, weight-count_{\mathcal{P}}(x_l, x_u, y_l, y_u) = \sum_{(x,y,w,\ell) \in \mathcal{R}} |\{j \in [1 ... n] : T^{\infty}[j - 7\tau ... j + 7\tau) \in T^{\infty}[j - 7\tau ... \ell + 7\tau)\}| = |\{j \in [1 ... n] : T^{\infty}[j - 7\tau ... j + 7\tau) \in Q\}| = |A|.

(b) Denote

$$A' = \{ s \in \mathsf{S} : x_l \preceq \overline{T^{\infty}[s - 7\tau \dots s)} \prec x_u \text{ and } y_l \preceq T^{\infty}[s \dots s + 7\tau) \prec y_u \}.$$

In the second step, we prove that A = A'.

- Let $i \in A$. Then, there exists $s \in S_{\text{comp}}$ such that $T^{\infty}[i 7\tau ...i + 7\tau) = T^{\infty}[s 7\tau ...s + 7\tau)$, $x_l \preceq T^{\infty}[s 7\tau ...s) \prec x_u$, and $y_l \preceq T^{\infty}[s ...s + 7\tau) \prec y_u$. Thus, we also have $x_l \preceq T^{\infty}[i 7\tau ...i) \prec x_u$ and $y_l \preceq T^{\infty}[i ...i + 7\tau) \prec y_u$. It therefore remains to show $i \in S$. First, note that by $s \in S_{\text{comp}} \subseteq S \subseteq [1 ...n 2\tau + 1]$, we have $s \in [1 ...n 2\tau + 1]$. Thus, by the uniqueness of T[n] in T and by $T^{\infty}[i ...i + 2\tau) = T^{\infty}[s ...s + 2\tau)$, we have $i \in [1 ...n 2\tau + 1]$. Consequently, by the consistency of S (Definition 2.4), we have $i \in S$. Thus, $i \in A'$.
- Let $s \in A'$. Then, $s \in S$, $x_l \preceq \overline{T^{\infty}[s-7\tau \ldots s)} \prec x_u$, and $y_l \preceq T^{\infty}[s \ldots s+7\tau) \prec y_u$. Denote $X = T^{\infty}[s-7\tau \ldots s+7\tau)$ and $i_{\text{left}} = \min\{i \in [1 \ldots n] : T^{\infty}[i-7\tau \ldots i+7\tau) = X\}$. We will prove that it holds $i_{\text{left}} \in S_{\text{comp}}$. First, note that $s+2\tau \leq n+1$ and $i_{\text{left}} \leq s$. Thus, $T[s \ldots s+2\tau) = T[i_{\text{left}} \ldots i_{\text{left}} + 2\tau)$. By the consistency of S (Definition 2.4), we thus have $i_{\text{left}} \in S$. To show $i_{\text{left}} \in S_{\text{comp}}$, we need to additionally prove that $i_{\text{left}} \in C(14\tau, T)$, i.e., that there exists $j \in [1 \ldots n]$ such that $j = \min \operatorname{Occ}_{14\tau}(j, T)$ and $i_{\text{left}} \in [j \ldots j+14\tau)$. We consider two cases.
 - If $i_{\text{left}} \leq 14\tau$, then it suffices to take j = 1.
 - Otherwise, let $j = i_{\text{left}} 7\tau$. By definition of i_{left} , for every $i \in [1 ... i_{\text{left}})$, we have $T^{\infty}[i 7\tau ... i + 7\tau) \neq X$. This implies that for every $i \in [1 ... i_{\text{left}} 7\tau)$, it holds $T^{\infty}[i ... i + 14\tau) \neq X$. Thus, by $T^{\infty}[j ... j + 14\tau) = X$, we obtain $j = \min \text{Occ}_{14\tau}(j, T)$ and hence $i_{\text{left}} \in [i_{\text{left}} 7\tau ... i_{\text{left}} + 7\tau) = [j ... j + 14\tau)$.

We have thus proved that there exists $\underline{i} \in \mathsf{S}_{\mathrm{comp}}$ such that $T^{\infty}[s-7\tau\mathinner{.\,.} s+7\tau) = T^{\infty}[i-7\tau\mathinner{.\,.} i+7\tau)$. By our assumption on s, it holds $x_l \preceq \overline{T^{\infty}[i-7\tau\mathinner{.\,.} i)} \prec x_u$ and $y_l \preceq T^{\infty}[i\mathinner{.\,.} i+7\tau) \prec y_u$. Thus, $s \in A$.

- (c) Combining (a) and (b), we obtain weight-count_P $(x_l, x_u, y_l, y_u) = |A| = |A'|$, i.e., the claim.
- 2. The proof as analogous, except we replace the condition $y_l \leq T^{\infty}[s ... s + 7\tau) \prec y_u$ in the definition of Q and A' with the condition $T^{\infty}[s ... s + 7\tau) \leq y_u$.
- 3. The proof is analogous, except we remove the condition $y_l \leq T^{\infty}[s ... s + 7\tau) \prec y_u$ in the definition of Q and A'.

Lemma 8.24. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$. Let S be a τ -synchronizing set of T and let $S_{\text{comp}} = \text{comp}_{14\tau}(S, T)$. Let $\mathcal{P} = \text{StrStrPoints}_{7\tau}(S_{\text{comp}}, T)$ (Definition 6.3). For any x_l, x_u, y_l, y_u such that weight-count $_{\mathcal{P}}(x_l, x_u, y_l, y_u) > 0$, it holds

$$\operatorname{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_u) = \min\{s \in \mathsf{S} : x_l \preceq \overline{T^{\infty}[s - 7\tau \dots s)} \prec x_u \text{ and } y_l \preceq T^{\infty}[s \dots s + 7\tau) \prec y_u\}.$$

Proof. The proof consists of two steps labeled (a) through (b).

- (a) Let Q and A be defined as in the proof of Lemma 8.23(1), i.e., $Q = \{T^{\infty}[s-7\tau ...s+7\tau) : s \in S_{\text{comp}}, x_l \preceq \overline{T^{\infty}[s-7\tau ...s)} \prec x_u \text{ and } y_l \preceq T^{\infty}[s...s+7\tau) \prec y_u\}$ and $A = \{i \in [1...n] : T^{\infty}[i-7\tau ...i+7\tau) \in Q\}$. Denote $\mathcal{R} = \{(x,y,w,\ell) \in \mathcal{P} : x_l \preceq x \prec x_u \text{ and } y_l \preceq y \prec y_u\}$. In the first step, we prove that $r\text{-min}_{\mathcal{P}}(x_l,x_u,y_l,y_u) = \min A$.
 - Let $i = \min A$. By definition of A, we then have $T^{\infty}[i 7\tau \dots i + 7\tau) \in Q$. Denote $S = T^{\infty}[i 7\tau \dots i + 7\tau)$. Recall from the proof of Lemma 8.23(1), that the function $g: \mathcal{R} \to Q$ defined such that for every $p = (x, y, w, \ell) \in \mathcal{R}$, $g(p) = T^{\infty}[\ell 7\tau \dots \ell + 7\tau)$, is a bijection. Let us thus consider $p_S = g^{-1}(S) \in \mathcal{R}$. Denote $p_S = (x_S, y_S, w_S, \ell_S)$. Observe, that we then have $T^{\infty}[\ell_S 7\tau \dots \ell_S + 7\tau) = g(p_S) = S = T^{\infty}[i 7\tau \dots i + 7\tau)$. By Definition 6.3, it then holds $\ell_S = \min\{i' \in [1 \dots n]: T^{\infty}[i' 7\tau \dots i' + 7\tau) = S\}$. Consequently, $\ell_S \leq i$ and thus $\operatorname{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_u) = \min_{(x, y, w, \ell) \in \mathcal{R}} \ell \leq \ell_S \leq i = \min A$.
 - Let $p_{\min} = (x_{\min}, y_{\min}, w_{\min}, \ell_{\min}) \in \mathcal{R}$ be such that $\operatorname{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_u) = \min_{(x, y, w, \ell) \in \mathcal{R}} \ell = \ell_{\min}$. Let $S = g(p_{\min}) = T^{\infty}[\ell_{\min} 7\tau \dots \ell_{\min} + 7\tau)$, where $g : \mathcal{R} \to Q$ is the bijection defined above. We then have $S \in Q$. Consequently, by $\ell_{\min} \in [1 \dots n]$ and $T^{\infty}[\ell_{\min} 7\tau \dots \ell_{\min} + 7\tau) = S$, we have $\ell_{\min} \in A$. Thus, $\min A \leq \ell_{\min} = \operatorname{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_u)$.
- (b) Let A' be defined as in the proof of Lemma 8.23(1), i.e., $A' = \{s \in \mathbb{S} : x_l \leq \overline{T^{\infty}[s 7\tau ..s)} \prec x_u \text{ and } y_l \leq T^{\infty}[s ..s + 7\tau) \prec y_u\}$. In the proof of Lemma 8.23(1), we showed that A = A'. Combining this with $\operatorname{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_u) = \min A$ shown above, we obtain $\operatorname{r-min}_{\mathcal{P}}(x_l, x_u, y_l, y_u) = \min A = \min A' = \min \{s \in \mathbb{S} : x_l \leq \overline{T^{\infty}[s 7\tau ..s)} \prec x_u \text{ and } y_l \leq \overline{T^{\infty}[s ..s + 7\tau)} \prec y_u\}$, i.e., the claim.

8.3.4 Basic Navigation Primitives

Lemma 8.25. Let $\tau \in [1..\lfloor \frac{n}{2} \rfloor]$ and $\ell, \ell', \ell' \in \mathbb{Z}_{>0}$ be such that $3\tau - 1 \le \ell \le \ell' \le \ell''$. Let S be a τ -synchronizing set of T and let $S_{\text{comp}} = \text{comp}_{\ell''}(S,T)$. Let $j \in [1..n - 3\tau + 2] \setminus R(\tau,T)$ and $j' \in \text{Occ}_{\ell}(j,T)$ be a position satisfying $j' = \min \text{Occ}_{\ell'}(j',T)$. Then, $\text{succ}_{S}(j) - j = \text{succ}_{S_{\text{comp}}}(j') - j'$.

Proof. We begin by observing that $j \in [1..n-3\tau+2]$ implies $3\tau-1 \le n$. It thus follows by Lemma 8.21 that $\mathsf{S} \ne \emptyset$ and $\max \mathsf{S} \ge n-3\tau+2$. Consequently, $\mathrm{succ}_{\mathsf{S}}(j)$ is well-defined. Next, we prove that $j' \in [1..n-3\tau+2]$ and $T[j..j+3\tau-1) = T[j'..j'+3\tau-1)$. By $3\tau-1 \le \ell$, we have $j' \in \mathrm{Occ}_{\ell}(j,T) \subseteq \mathrm{Occ}_{3\tau-1}(j,T)$. Consider two cases:

- If j = j', then the claim follows immediately.
- If $j \neq j'$, then by Lemma 8.11, $j' \in \text{Occ}_{3\tau-1}(j,T)$ implies $\text{LCE}_T(j,j') \geq 3\tau 1$, i.e., $j' \in [1 ... n 3\tau + 2]$ and $T[j ... j + 3\tau 1) = T[j' ... j' + 3\tau 1)$.

In particular, we have $j' \in [1 ... n - 3\tau + 2]$. By Lemma 8.21, this implies $\mathsf{S}_{\text{comp}} \neq \emptyset$ and $\max \mathsf{S}_{\text{comp}} \geq n - 3\tau + 2$. Hence, $\mathrm{succ}_{\mathsf{S}_{\text{comp}}}(j')$ is well-defined.

We show the main claim in two steps:

- 1. By the consistency of S (Definition 2.4), it follows that for every $t \in [0...\tau)$, $j' + t \in S$ holds if and only if $j + t \in S$. Observe, however, that by the density condition (Definition 2.4), we have $S \cap [j...j + \tau) \neq \emptyset$. Consequently, letting $s_j = \text{succ}_S(j)$, it holds $s_j j < \tau$. By the above property for $t \in [0...s_j j]$ we thus obtain that $s_{j'} j' = s_j j$, where $s_{j'} = \text{succ}_S(j')$.
- 2. Denote $s' = \operatorname{succ}_{S_{\operatorname{comp}}}(j')$. Observe that $j' = \min \operatorname{Occ}_{\ell''}(j',T)$. Otherwise, there would exist $j'' \in [1 \dots j')$ such that $T^{\infty}[j'' \dots j'' + \ell'') = T^{\infty}[j' \dots j' + \ell'')$. By $\ell' \leq \ell''$, we would then have $T^{\infty}[j'' \dots j'' + \ell') = T^{\infty}[j' \dots j' + \ell')$, which would contradict $j' = \min \operatorname{Occ}_{\ell'}(j',T)$. Recalling from Definition 4.10 that $S_{\operatorname{comp}} = S \cap C(\ell'',T)$, it follows that $[j' \dots j' + \ell'') \subseteq C(\ell'',T)$, and thus for every $q \in [0 \dots \ell'')$, $j' + q \in S$ holds if and only if $j' + q \in S_{\operatorname{comp}}$. Applying this for $q \in [0 \dots s_{j'} j']$ and recalling that $s_{j'} j' < \tau$, we obtain $\operatorname{succ}_{S_{\operatorname{comp}}}(j') j' = \operatorname{succ}_{S}(j') j'$, i.e., $s' j' = s_{j'} j'$. Combining this with $s_j j = s_{j'} j'$, we thus have $s_j j = s_{j'} j' = s' j'$.

Proposition 8.26. Let $k \in [4 .. \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $j \in [1 ... n - 3\tau + 2] \backslash R(\tau, T)$. Given the structure CompSANonperiodic(T), the value k, the position j, and any $j' \in \operatorname{Occ}_{\ell}(j, T)$ satisfying $j' = \min \operatorname{Occ}_{2\ell}(j', T)$ as input, we can compute $\operatorname{succ}_{S_k}(j)$ in $\mathcal{O}(\log n)$ time.

Proof. First, using binary search over the array $A_{\text{comp},k}$, we compute $s' = \text{succ}_{S_{\text{comp},k}}(j')$ in $\mathcal{O}(\log n)$ time. Observe now that by $\ell \geq 16$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$, it holds $2\ell \leq 14\tau$. By Lemma 8.25 (with $\ell = \ell$, $\ell' = 2\ell$, and $\ell'' = 14\tau$), we then obtain $\text{succ}_{S_k}(j) = j + (s' - j')$.

8.3.5 Computing the Size of Posbeg and Posend

Combinatorial Properties

Lemma 8.27. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let S be a τ -synchronizing set of T. Let $P \in \Sigma^m$ be a τ -nonperiodic pattern such that $m \geq 3\tau - 1$, no nonempty suffix of T is a proper prefix of P, and P is prefixed with some $D \in \mathcal{D}(\tau, T, S)$ (Definition 8.19). Let $c = \max \Sigma$ and x, x', y, y' be such that:

- \overline{x} is a prefix of D and $|\overline{x}| + 2\tau = |D|$,
- $x' = xc^{\infty}$,
- $\overline{x}y = P[1 .. \min(m, \ell)],$
- $\overline{x}y' = P[1 .. \min(m, 2\ell)].$

Then, it holds:

 $\begin{array}{l} \text{1. } \operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T) = \{s - |x| : s \in \mathsf{S}, \ x \preceq \overline{T^{\infty}[s - 7\tau \ldots s)} \prec x', \ and \ y \preceq T^{\infty}[s \ldots s + 7\tau) \prec y'\}, \\ \text{2. } \operatorname{Pos}^{\operatorname{end}}_{\ell}(P,T) = \{s - |x| : s \in \mathsf{S}, \ x \preceq \overline{T^{\infty}[s - 7\tau \ldots s)} \prec x', \ and \ y \preceq_{\operatorname{inv}} T^{\infty}[s \ldots s + 7\tau) \prec_{\operatorname{inv}} y'\}. \end{array}$

Proof. 1. Let $j \in \operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T)$. Denote s = j + |x| We need to prove (a) $s \in S$, (b) $x \preceq \overline{T^{\infty}[s - 7\tau \dots s)} \prec x'$, and (c) $y \preceq T^{\infty}[s \dots s + 7\tau) \prec y'$.

- (a) By Lemma 8.4, $T[j ...n] \prec P$ and $lcp(P, T[j ...n]) \in [\ell ...2\ell)$. On the other hand, by $D \in \mathcal{D}(\tau, T, S)$ there exists $j' \in [1 ...n 3\tau + 2] \setminus R(\tau, T)$ such that $j' \in Occ(D, T)$ and $succ_S(j') = j' + |D| 2\tau = j' + |x|$. Thus, $j' + |x| \in S$. Since $j' \in [1 ...n 3\tau + 2] \setminus R(\tau, T)$ implies $[j' ...j' + \tau) \cap S \neq \emptyset$, we thus obtain $|x| < \tau$ and hence $|D| = |x| + 2\tau \le 3\tau 1 \le \ell$. By $lcp(P, T[j ...n]) \ge \ell$ and D being a prefix of P, we therefore have $j \in Occ(D, T)$. Consequently, $T[j' + |x| ...j' + |x| + 2\tau) = T[j + |x| ...j + |x| + 2\tau)$, and hence by the consistency of S, it holds $s = j + |x| \in S$.
- (b) By the above, $j \in \text{Occ}(D, T)$. Since \overline{x} is a prefix of D, we thus have $T[j ... j + |x|) = \overline{x}$, or equivalently, $\overline{T[j ... j + |x|)} = x$, which we can write as $\overline{T[s |x| ... s)} = x$. Since $|x| < \tau$, this implies that x is a prefix of $T^{\infty}[s 7\tau ... s)$. By $x' = xc^{\infty}$, we thus have $x \leq \overline{T^{\infty}[s 7\tau ... s)} \prec x'$.
- (c) Above we observed that \overline{x} is a prefix of both P and T[j ...n]. Thus, recalling that j + |x| = s and letting P' be such that $\overline{x}P' = P$, the conditions $T[j ...n] \prec P$ and $\operatorname{lcp}(P, T[j ...n]) \in [\ell ... 2\ell)$ imply $T[s ...n] \prec P'$ and $\operatorname{lcp}(P', T[s ...n]) \in [\ell |x| ... 2\ell |x|)$. Denote m' = |P'| = m |x|. By Lemma 8.8(3) with $\ell_1 = \ell |x|$ and $\ell_2 = 2\ell |x|$, it holds $P_1 \preceq T[s ...n] \prec P_2$, where $P_1 = P'[1 ... \min(m', \ell_1)] = P'[1 ... \min(m, \ell) |x|]$ and $P_2 = P'[1 ... \min(m', \ell_2)] = P'[1 ... \min(m, 2\ell) |x|]$. Note that $\ell_1 < \ell_2 \le 7\tau$. By Lemma 8.9(3), we thus obtain $P_1 \preceq T^{\infty}[s ... s + 7\tau) \prec P_2$. It remains to observe that $\overline{x}P_1$ and $\overline{x}P_2$ are prefixes of P, and it holds $|\overline{x}P_1| = |\overline{x}| + \min(m, \ell) |x| = \min(m, \ell)$ and $|\overline{x}P_2| = |\overline{x}| + \min(m, 2\ell) |x| = \min(m, 2\ell)$. Thus, $P_1 = y$ and $P_2 = y'$, and hence we obtain $y \preceq T^{\infty}[s ... s + 7\tau) \prec y'$.

Let us now consider any $s \in S$ satisfying $x \leq \overline{T^{\infty}[s-7\tau..s]} \prec x'$ and $y \leq T^{\infty}[s..s+7\tau) \prec y'$. Denote j = s - |x|. We will prove that $j \in \operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T)$. By Lemma 8.4, this is equivalent to (a) $j \in [1..n]$ and (b) $T[j..n] \prec P$ and $\operatorname{lcp}(P,T[j..n]) \geq [\ell..2\ell)$.

- (a) Recall that $x' = xc^{\infty}$. The inequality $x \leq \overline{T^{\infty}[s 7\tau ... s)} \prec x'$ thus implies that x is a prefix of $\overline{T^{\infty}[s 7\tau ... s)}$, or equivalently, $T^{\infty}[s |x| ... s) = \overline{x}$. Suppose that s |x| < 1. Then, there exists a nonempty prefix of \overline{x} that is a suffix of T. Note, however, that $|x| < |D| \leq m$, i.e., \overline{x} is a proper prefix of P. Thus, this contradicts the assumption about T not having any nonempty suffix that is a proper prefix of P. We thus have $s |x| \geq 1$, and hence $j \in [1...n]$.
- (b) Recall that $|y| = \min(m, \ell) |x| \le \ell$ and $|y'| = \min(m, 2\ell) |x| \le 2\ell$. By $2\ell \le 7\tau$ and Lemma 8.9(3), we thus obtain $y \le T[s \dots n] \prec y'$. Next, note that letting P' be such that $\overline{x}P' = P$, it holds $y = P'[1 \dots \min(m', \ell_1)]$ and $y' = P'[1 \dots \min(m, \ell_2)]$, where m' = |P'| = m |x|, $\ell_1 = \ell |x|$ and $\ell_2 = 2\ell |x|$. By $\ell_1 < \ell_2$ and Lemma 8.8(3), we thus obtain $T[s \dots n] \prec P'$ and $\operatorname{lcp}(P', T[s \dots n]) \in [\ell_1 \dots \ell_2) = [\ell |x| \dots 2\ell |x|)$. It remains to note that above we observed that $T[s |x| \dots s) = \overline{x}$. Thus, we obtain $T[j \dots n] = T[s |x| \dots n] \prec \overline{x}P' = P$ and $\operatorname{lcp}(P, T[j \dots n]) = \operatorname{lcp}(\overline{x}P', \overline{x}T[s \dots n]) = |x| + \operatorname{lcp}(P', T[s \dots n]) \in [\ell \dots 2\ell)$.
- 2. The proof follows by first applying Lemma 8.4 and then symmetrically repeating all arguments above.

Lemma 8.28. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let S be a τ -synchronizing set of T. Let $j \in [1..n-3\tau+2] \setminus R(\tau,T)$, $c = \max \Sigma$, $s = \operatorname{succ}_{S}(j)$, and x, x', y, y' be such that:

- $\overline{x} = T[j \dots s),$
- $x' = xc^{\infty}$,
- $\overline{x}y = T[j ... \min(n+1, j+\ell)),$
- $\overline{x}y' = T[j ... \min(n+1, j+2\ell)).$

Then:

$$\begin{array}{l} \text{1. } \operatorname{Pos}^{\operatorname{beg}}_{\ell}(j,T) = \{s - |x| : s \in \mathsf{S}, \ x \preceq \overline{T^{\infty}[s - 7\tau \mathinner{\ldotp\ldotp} s)} \prec x', \ and \ y \preceq T^{\infty}[s \mathinner{\ldotp\ldotp} s + 7\tau) \prec y'\}, \\ \text{2. } \operatorname{Pos}^{\operatorname{end}}_{\ell}(j,T) = \{s - |x| : s \in \mathsf{S}, \ x \preceq \overline{T^{\infty}[s - 7\tau \mathinner{\ldotp\ldotp} s)} \prec x', \ and \ y \preceq_{\operatorname{inv}} T^{\infty}[s \mathinner{\ldotp\ldotp} s + 7\tau) \prec_{\operatorname{inv}} y'\}. \end{array}$$

Proof. 1. Denote P = T[j ...n] and m = |P| = n - j + 1. Note that $j \in [1 ...n - 3\tau + 2]$ and $j \notin R(\tau, T)$ imply that $m \ge 3\tau - 1$ and $\operatorname{per}(P[1 ...3\tau - 1]) > \frac{1}{3}\tau$. Thus, P is τ -nonperiodic. Next, observe that since T[n] does not occur in T[1 ...n), no nonempty suffix of T is a proper prefix of P. Next, note that by $j \in [1 ...n - 3\tau + 2] \setminus R(\tau, T)$, it holds $[j ...j + \tau) \cap S \ne \emptyset$. Thus, $s - j < \tau$, and hence letting $D = T[j ...s + 2\tau)$ it holds $D \in \mathcal{D}(\tau, T, S)$. Thus, P has D as a prefix. Observe that \overline{x} is a prefix of D and it holds

 $|\overline{x}|+2\tau=|D|. \text{ Finally, observe that it holds } \min(n+1,j+\ell)-j=\min(n+1-j,\ell)=\min(m,\ell) \text{ and } \min(n+1,j+2\ell)-j=\min(n+1-j,2\ell)=\min(m,2\ell). \text{ Thus, } \overline{x}y=T[j\ldots\min(n+1,j+\ell))=P[1\ldots\min(m,\ell)] \text{ and } \overline{x}y'=T[j\ldots\min(n+1,j+2\ell))=P[1\ldots\min(m,2\ell)]. \text{ We have thus proved that all assumptions of Lemma 8.27 hold for } P. \text{ The claim thus follows by combining the definition of } \operatorname{Pos}_{\ell}^{\operatorname{beg}}(j,T) \text{ and Lemma 8.27}(1), \text{ i.e., } \operatorname{Pos}_{\ell}^{\operatorname{beg}}(j,T)=\operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T)=\{s-|x|:s\in \mathsf{S},\ x\preceq \overline{T^{\infty}[s-7\tau\ldots s)}\prec x',\ \text{ and } y\preceq T^{\infty}[s\ldots s+7\tau)\prec y'\}.$ 2. The proof proceeds symmetrically, except we utilize Lemma 8.27(2).

Lemma 8.29. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let S be a τ -synchronizing set of T and $S_{\text{comp}} = \text{comp}_{14\tau}(S, T)$. Let $j \in [1..n - 3\tau + 2] \setminus R(\tau, T)$, $c = \max \Sigma$, $s = \text{succ}_S(j)$, and x, x', y, y' be such that:

- $\overline{x} = T[j \dots s),$
- $x' = x\tilde{c}^{\infty}$,
- $\overline{x}y = T[j ... \min(n+1, j+\ell)),$
- $\overline{x}y' = T[j ... \min(n+1, j+2\ell)).$

 $Then, \ |\operatorname{Pos}_{\ell}^{\operatorname{beg}}(j,T)| = \operatorname{weight-count}_{\mathcal{P}}(x,x',y,y'), \ where \ \mathcal{P} = \operatorname{StrStrPoints}_{7\tau}(\mathsf{S}_{\operatorname{comp}},T) \ (Definition \ \textbf{6.3}).$

Proof. Increasing all elements of a set by the same value does change its cardinality. By Lemma 8.28(1) and Lemma 8.23, we thus obtain

$$\begin{split} |\operatorname{Pos}^{\operatorname{beg}}_{\ell}(j,T)| &= |\{s - |x| : s \in \mathsf{S}, \ x \preceq \overline{T^{\infty}[s - 7\tau \ldots s)} \prec x' \text{ and } y \preceq T^{\infty}[s \ldots s + 7\tau) \prec y'\}| \\ &= |\{s \in \mathsf{S} : x \preceq \overline{T^{\infty}[s - 7\tau \ldots s)} \prec x' \text{ and } y \preceq T^{\infty}[s \ldots s + 7\tau) \prec y'\}| \\ &= \mathsf{weight\text{-}count}_{\mathcal{P}}(x, x', y, y'). \end{split}$$

Query Algorithms

Proposition 8.30. Let $k \in [4 .. \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $j \in [1 ... n] \setminus \mathsf{R}(\tau, T)$. Given the structure CompSANonperiodic(T), the value k, the position j, and any $j' \in \mathrm{Occ}_{\ell}(j,T)$ satisfying $j' = \min \mathrm{Occ}_{2\ell}(j',T)$ as input, we can compute $|\mathrm{Pos}^{\mathrm{beg}}_{\ell}(j,T)|$ in $\mathcal{O}(\log^{2+\epsilon} n)$ time.

Proof. If $j > n - 3\tau + 2$ then by $3\tau - 1 \le \ell$ and the uniqueness of T[n] in T, it holds $|\operatorname{Occ}_{\ell}(j,T)| = 1$. By $|\operatorname{Occ}_{2\ell}(j,T)| > 0$ and $\operatorname{Occ}_{2\ell}(j,T) \subseteq \operatorname{Occ}_{\ell}(j,T)$ we return $|\operatorname{Pos}_{\ell}^{\operatorname{beg}}(j,T)| = 0$.

Let us now assume $j \in [1..n - 3\tau + 2]$. Denote $\mathcal{P} = \text{StrStrPoints}_{7\tau}(\mathsf{S}_{\text{comp},k},T)$. The algorithm consists of two steps:

- 1. First, using Proposition 8.26, we compute $s = \operatorname{succ}_{S_k}(j)$ in $\mathcal{O}(\log n)$ time.
- 2. Let x, x', y, y' be such that $T[j ... s) = \overline{x}, x' = xc^{\infty}, \overline{x}y = T[j .. \min(n+1, j+\ell))$, and $\overline{x}y' = T[j .. \min(n+1, j+2\ell))$. By Lemma 8.29, $|\operatorname{Pos}_{\ell}^{\operatorname{beg}}(j,T)| = \operatorname{weight-count}_{\mathcal{P}}(x, x', y, y') = \operatorname{weight-count}_{\mathcal{P}}(x, x', y') \operatorname{weight-count}_{\mathcal{P}}(x, x', y)$ (where the second equality holds by $y \leq y'$). Observe also that using the structure $\operatorname{CompSANonperiodic}(T)$, we can perform LCE_T and $\operatorname{LCE}_{\overline{T}}$ queries in $\mathcal{O}(\log n)$ time, and we can access any symbol of T in $\mathcal{O}(\log n)$ time. Thus, we can compare any two substrings of T^{∞} or $\overline{T^{\infty}}$ (specified with their starting positions and lengths) in $t_{\operatorname{cmp}} = \mathcal{O}(\log n)$ time. Thus, using the structure from Proposition 6.4 for $q = 7\tau$ and $P = \mathsf{S}_{\operatorname{comp},k}$ (which is a component of $\operatorname{CompSANonperiodic}(T)$) first with parameters $(i, q_l, q_r) = (s, s-j, \min(n+1, j+2\ell) s)$ and then $(i, q_l, q_r) = (s, s-j, \min(n+1, j+\ell) s)$ we compute $|\operatorname{Pos}_{\ell}^{\operatorname{beg}}(j, T)|$ in $\mathcal{O}(\log^{2+\epsilon} n + t_{\operatorname{cmp}} \log n) = \mathcal{O}(\log^{2+\epsilon} n)$ time.

In total, we spend $\mathcal{O}(\log^{2+\epsilon} n)$ time.

8.3.6 Computing the Size of Occ

Combinatorial Properties

Lemma 8.31. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let S be a τ -synchronizing set of T. Let $P \in \Sigma^m$ be a τ -nonperiodic pattern such that $m \geq 3\tau - 1$, no nonempty suffix of T is a proper prefix of P, and P is prefixed with some $D \in \mathcal{D}(\tau, T, S)$ (Definition 8.19). Let $d \in [\ell ... 2\ell]$, $c = \max \Sigma$, and x, x', y, y' be such that:

- \overline{x} is a prefix of D and $|\overline{x}| + 2\tau = |D|$,
- $x' = xc^{\infty}$,
- $\overline{x}y = P[1 .. \min(m, d)],$
- $y' = yc^{\infty}$.

Then, $\operatorname{Occ}_d(P,T) = \{s - |x| : s \in S, \ x \leq \overline{T^{\infty}[s - 7\tau \dots s)} \prec x', \ and \ y \leq T^{\infty}[s \dots s + 7\tau) \prec y'\}.$

Proof. Let $j \in \text{Occ}_d(P,T)$. Denote s = j + |x| We need to prove (a) $s \in S$, (b) $x \preceq \overline{T^{\infty}[s - 7\tau ...s)} \prec x'$, and (c) $y \preceq T^{\infty}[s ...s + 7\tau) \prec y'$.

- (a) By Definition 2.1, it holds $j \in [1 ...n]$ and $lcp(P, T[j ...n]) \ge \min(m, d)$. By $D \in \mathcal{D}(\tau, T, \mathsf{S})$ there exists $j' \in [1 ...n 3\tau + 2] \setminus \mathsf{R}(\tau, T)$ such that $j' \in \mathsf{Occ}(D, T)$ and $\mathsf{succ}_{\mathsf{S}}(j') = j' + |D| 2\tau = j' + |x|$. Thus, $j' + |x| \in \mathsf{S}$. Since $j' \in [1 ...n 3\tau + 2] \setminus \mathsf{R}(\tau, T)$ implies $[j' ...j' + \tau) \cap \mathsf{S} \ne \emptyset$, we thus obtain $|x| < \tau$ and hence $|D| = |x| + 2\tau \le 3\tau 1 \le \ell$. By $lcp(P, T[j ...n]) \ge \min(m, d) \ge 3\tau 1$ and by D being a prefix of P, we therefore have $j \in \mathsf{Occ}(D, T)$. Consequently, $T[j' + |x| ...j' + |x| + 2\tau) = T[j + |x| ...j + |x| + 2\tau)$, and hence by the consistency of S , it holds $s = j + |x| \in \mathsf{S}$.
- (b) By the above, $j \in \text{Occ}(D, T)$. Since \overline{x} is a prefix of D, we thus have $T[j ... j + |x|) = \overline{x}$, or equivalently, $\overline{T[j ... j + |x|)} = x$, which we can write as $\overline{T[s |x| ... s)} = x$. Since $|x| < \tau$, this implies that x is a prefix of $T^{\infty}[s 7\tau ... s)$. By $x' = xc^{\infty}$, we thus have $x \leq T^{\infty}[s 7\tau ... s) < x'$.
- (c) Recall that $\overline{x}y = P[1 ... \min(m, d)]$. On the other hand, we assumed $lcp(P, T[j ... n]) \ge \min(m, d)$. Thus, $\overline{x}y$ is a prefix of T[j ... n]. In particular, y is a prefix of T[s ... n]. Since $|y| \le d \le 2\ell \le 7\tau$, we thus obtain that y if a prefix of $T^{\infty}[s ... s + 7\tau)$. By $y' = yc^{\infty}$, we thus have $y \le T^{\infty}[s ... s + 7\tau) < y'$.

Let us now consider any $s \in S$ satisfying $x \leq \overline{T^{\infty}[s-7\tau..s)} \prec x'$ and $y \leq T^{\infty}[s..s+7\tau) \prec y'$. Denote j = s - |x|. We will prove that $j \in \operatorname{Occ}_d(P,T)$, i.e., (a) $j \in [1..n]$ and (b) $\operatorname{lcp}(P,T[j..n]) \geq \min(m,d)$.

- (a) Recall that $x' = xc^{\infty}$. The inequality $x \leq \overline{T^{\infty}[s 7\tau ...s)} \prec x'$ thus implies that x is a prefix of $\overline{T^{\infty}[s 7\tau ...s)}$, or equivalently, $T^{\infty}[s |x| ...s) = \overline{x}$. Suppose that s |x| < 1. Then, there exists a nonempty prefix of \overline{x} that is a suffix of T. Note, however, that $|x| < |D| \leq m$, i.e., \overline{x} is a proper prefix of P. Thus, this contradicts the assumption about T not having any nonempty suffix that is a proper prefix of P. We thus have $s |x| \geq 1$, and hence $j \in [1...n]$.
- (b) Recall that $y' = yc^{\infty}$. The assumption $y \leq T^{\infty}[s \ldots s + 7\tau) \prec y'$ thus implies that y is a prefix of $T^{\infty}[s \ldots s + 7\tau)$. Combining with the above, we thus obtain that $\overline{x}y$ is a prefix of $T^{\infty}[j \ldots s + 7\tau)$, i.e., $T^{\infty}[j \ldots j + \min(m,d)) = \overline{x}y$. Suppose that $j + \min(m,d) > n+1$. Then, there exists a nonempty suffix of T that is a proper prefix of $\overline{x}y$. But since $\overline{x}y$ is a prefix of P, this contradicts the assumption about T not having a nonempty suffix that is a proper prefix of P. Thus, we have $j + \min(n,d) \leq n+1$, and hence $\overline{x}y = T[j \ldots j + \min(m,d))$. By $\overline{x}y = P[1 \ldots \min(m,d)]$, this implies $lcp(P,T[j \ldots n]) \geq |\overline{x}y| = \min(m,d)$.

Lemma 8.32. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let S be a τ -synchronizing set of T. Let $j \in [1..n-3\tau+2] \setminus R(\tau,T)$, $c = \max \Sigma$, $s = \operatorname{succ}_{S}(j)$, $d \in [\ell..2\ell]$, and x, x', y, y' be such that:

- $\overline{x} = T[j ... s),$
- $x' = xc^{\infty}$,
- $\overline{x}y = T[j ... \min(n+1, j+d)),$
- $y' = yc^{\infty}$.

Then, $\operatorname{Occ}_d(j,T) = \{s - |x| : s \in S, \ x \leq \overline{T^{\infty}[s - 7\tau \dots s]} \prec x', \ and \ y \leq T^{\infty}[s \dots s + 7\tau) \prec y'\}.$

Proof. Denote P = T[j ...n] and m = |P| = n - j + 1. Note that $j \in [1...n - 3\tau + 2]$ and $j \notin R(\tau, T)$ imply that $m \geq 3\tau - 1$ and $\operatorname{per}(P[1..3\tau - 1]) > \frac{1}{3}\tau$. Thus, P is τ -nonperiodic. Next, observe that since T[n] does not occur in T[1...n), no nonempty suffix of T is a proper prefix of P. Next, note that by $j \in [1...n - 3\tau + 2] \setminus R(\tau, T)$, it holds $[j...j + \tau) \cap S \neq \emptyset$. Thus, $s - j < \tau$, and hence letting $D = T[j...s + 2\tau)$ it holds $D \in \mathcal{D}(\tau, T, S)$. Thus, P has D as a prefix. Observe that \overline{x} is a prefix of D and it holds $|\overline{x}| + 2\tau = |D|$. Finally, observe that it holds $\min(n+1,j+d) - j = \min(n+1-j,d) = \min(m,d)$. Thus, $\overline{x}y = T[j...\min(n+1,j+d)) = P[1...\min(m,d)]$. We have thus proved that all assumptions of

Lemma 8.31 hold for P. The claim thus follows by combining the definition of $Occ_d(j,T)$ and Lemma 8.31, i.e., $\operatorname{Occ}_d(j,T) = \operatorname{Occ}_d(P,T) = \{s - |x| : s \in S, \ x \leq \overline{T^{\infty}[s - 7\tau ...s]} \prec x', \text{ and } y \leq T^{\infty}[s ...s + 7\tau) \prec y'\}.$

Lemma 8.33. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let S be a τ -synchronizing set of T and $S_{\text{comp}} = \text{comp}_{14\tau}(S, T)$. Let $j \in [1 ... n - 3\tau + 2] \setminus R(\tau, T)$, $c = \max \Sigma$, $s = \operatorname{succ}_{S}(j)$, $d \in [\ell ... 2\ell]$, and x, x', y, y' be such that

- $\overline{x} = T[j ... s),$
- $\bullet \ x' = xc^{\infty},$
- $\bullet \ \overline{x}y = T[j] \cdot \min(n+1, j+d)),$ $\bullet \ y' = yc^{\infty}.$

Then, $|\operatorname{Occ}_d(j,T)| = \operatorname{weight-count}_{\mathcal{P}}(x,x',y,y')$, where $\mathcal{P} = \operatorname{StrStrPoints}_{7\tau}(\mathsf{S}_{\operatorname{comp}},T)$ (Definition 6.3).

Proof. Increasing all elements of a set by the same value does change its cardinality. By Lemma 8.32 and Lemma 8.23, we thus obtain:

$$\begin{aligned} |\mathrm{Occ}_d(j,T)| &= |\{s-|x|: s \in \mathsf{S}, \ x \preceq \overline{T^\infty[s-7\tau \mathinner{\ldotp\ldotp} s)} \prec x' \text{ and } y \preceq T^\infty[s\mathinner{\ldotp\ldotp} s+7\tau) \prec y'\}| \\ &= |\{s \in \mathsf{S}: x \preceq \overline{T^\infty[s-7\tau \mathinner{\ldotp\ldotp} s)} \prec x' \text{ and } y \preceq T^\infty[s\mathinner{\ldotp\ldotp} s+7\tau) \prec y'\}| \\ &= \mathsf{weight\text{-}count}_{\mathcal{D}}(x,x',y,y'). \end{aligned} \quad \Box$$

Query Algorithms

Proposition 8.34. Let $k \in [4..\lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $j \in [1..n] \setminus \mathsf{R}(\tau,T)$. Given the structure CompSANonperiodic(T), the value k, the position j, and any $j' \in \text{Occ}_{\ell}(j,T)$ satisfying $j' = \min \text{Occ}_{2\ell}(j',T)$ as input, we can compute $|\operatorname{Occ}_{2\ell}(j,T)|$ in $\mathcal{O}(\log^{2+\epsilon} n)$ time.

Proof. If $j > n - 3\tau + 2$ then by $3\tau - 1 \le \ell$ and the uniqueness of T[n] in T, it holds $|\operatorname{Occ}_{\ell}(j,T)| = 1$. By $|\operatorname{Occ}_{2\ell}(j,T)| > 0$ and $\operatorname{Occ}_{2\ell}(j,T) \subseteq \operatorname{Occ}_{\ell}(j,T)$ we return $|\operatorname{Occ}_{2\ell}(j,T)| = 1$.

Let us now assume $j \in [1..n - 3\tau + 2]$. Denote $\mathcal{P} = \text{StrStrPoints}_{7\tau}(\mathsf{S}_{\text{comp},k},T)$. The algorithm consists of two steps:

- 1. First, using Proposition 8.26, we compute $s = \operatorname{succ}_{S_k}(j)$ in $\mathcal{O}(\log n)$ time.
- 2. Let x, x', y, y' be such that $T[j ... s) = \overline{x}$, $x' = xc^{\infty}$, $\overline{x}y = T[j ... \min(n+1, j+2\ell))$, and $y' = yc^{\infty}$. By Lemma 8.33, it then holds $|\operatorname{Occ}_{2\ell}(j,T)| = \mathsf{weight\text{-}count}_{\mathcal{P}}(x,x',y,y') = \mathsf{weight\text{-}count}_{\mathcal{P}}(x,x',y')$ weight-count_P(x, x', y) (where the last equality follows by $y \leq y'$). Observe also that using the structure CompSANonperiodic(T), we can perform LCE_T and LCE_{\overline{T}} queries in $\mathcal{O}(\log n)$ time, and we can access any symbol of T in $\mathcal{O}(\log n)$ time. Thus, we can compare any two substrings of T^{∞} or $\overline{T^{\infty}}$ (specified with their starting positions and lengths) in $t_{\text{cmp}} = \mathcal{O}(\log n)$ time. Thus, using the structure from Proposition 6.4 for $q = 7\tau$ and $P = S_{\text{comp},k}$ (which is a component of CompSANonperiodic(T)) with parameters $(i, q_l, q_r) = (s, s - j, \min(n + 1, j + 2\ell) - s)$ we compute $|\operatorname{Occ}_{2\ell}(j,T)|$ in $\mathcal{O}(\log^{2+\epsilon} n + t_{\text{cmp}} \log n) = \mathcal{O}(\log^{2+\epsilon} n)$ time.

In total, we spend $\mathcal{O}(\log^{2+\epsilon} n)$ time.

8.3.7 Computing a Position in Occ

Combinatorial Properties

Lemma 8.35. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let S be a τ -synchronizing set of T and $\mathsf{S}_{\text{comp}} = \mathsf{comp}_{14\tau}(\mathsf{S},T)$. Let $P \in \Sigma^m$ be a τ -nonperiodic pattern such that $m \geq 3\tau - 1$, no nonempty suffix of T is a proper prefix of P, and P is prefixed with $D \in \mathcal{D}(\tau, T, S)$ (Definition 8.19). Let $c = \max \Sigma$ and x, x', y, y' be such that:

- \overline{x} is a prefix of D and $|\overline{x}| + 2\tau = |D|$,
- $x' = xc^{\infty}$,
- $\overline{x}y = P[1 .. \min(m, \ell)].$

Let also $\mathcal{P} = \operatorname{Str}\operatorname{Str}\operatorname{Points}_{7\tau}(\mathsf{S}_{\operatorname{comp}},T)$ (Definition 6.3), $m = \operatorname{weight-count}_{\mathcal{P}}(x,x',y)$, and $b = \operatorname{RangeBeg}_{\ell}(P,T)$. Finally, let $i \in [1..n]$ be such that $\operatorname{SA}[i] \in \operatorname{Occ}_{2\ell}(P,T)$. Then:

- It holds $m + (i b) \in [1 ... \text{ weight-count}_{\mathcal{P}}(x, x')].$
- Every $p \in \text{weight-select}_{\mathcal{P}}(x, x', m + (i b))$ satisfies $p |x| \in \text{Occ}_{2\ell}(P, T)$.

Proof. The proof consists of three steps.

- 1. First, we prove that it holds $m+(i-b)\in [1...\text{weight-count}_{\mathcal{P}}(x,x')]$, i.e., the first claim. First, observe that $\mathrm{SA}[i]\in \mathrm{Occ}_2(P,T)\subseteq \mathrm{Occ}_\ell(P,T)$. Thus, i>b, which by $m\geq 0$ implies that $m+(i-b)\geq 1$. Denote $m'=\mathrm{weight-count}_{\mathcal{P}}(x,x')$. To prove $m+(i-b)\leq m'$, note that $i\leq \mathrm{RangeEnd}_\ell(P,T)$, i.e., letting $d=|\mathrm{Occ}_\ell(P,T)|$, we have $i-b\leq d$. Thus, it suffices to prove $m+d\leq m'$.
 - First, we observe that by Lemma 8.23(1), letting $S_{low} = \{s \in S : x \leq \overline{T^{\infty}[s 7\tau ...s)} \prec x' \text{ and } T^{\infty}[s ...s + 7\tau) \prec y\}$, it holds $m = \mathsf{weight\text{-}count}_{\mathcal{P}}(x, x', y) = |S_{low}|$.
 - By Lemma 8.31, it holds $\operatorname{Occ}_{\ell}(P,T) = \{s |x| : s \in S, \ x \leq \overline{T^{\infty}[s 7\tau \ldots s)} \prec x', \text{ and } y \leq T^{\infty}[s \ldots s + 7\tau) \prec yc^{\infty}\}$. Thus, letting $\mathsf{S}_{\operatorname{mid}} = \{s \in \mathsf{S} : x \leq \overline{T^{\infty}[s 7\tau \ldots s)} \prec x' \text{ and } y \leq T^{\infty}[s \ldots s + 7\tau) \prec yc^{\infty}\}$, we have $|\mathsf{S}_{\operatorname{mid}}| = |\operatorname{Occ}_{\ell}(P,T)| = d$.
 - Lastly, by Lemma 8.23(3), letting $S_{\text{all}} = \{s \in S : x \leq \overline{T^{\infty}[s 7\tau ...s)} \prec x'\}$, we have $m' = |S_{\text{all}}|$.

Note now that $\mathsf{S}_{\mathrm{low}} \subseteq \mathsf{S}_{\mathrm{all}}, \, \mathsf{S}_{\mathrm{mid}} \subseteq \mathsf{S}_{\mathrm{all}}, \, \mathrm{and} \, \mathsf{S}_{\mathrm{low}} \cap \mathsf{S}_{\mathrm{mid}} = \emptyset. \, \text{Thus, } m+d=|\mathsf{S}_{\mathrm{low}}|+|\mathsf{S}_{\mathrm{mid}}| \leq |\mathsf{S}_{\mathrm{all}}|=m'.$

- 2. In the second step, we show that for every $k \in [1 ... d]$, any $p \in \text{weight-select}_{\mathcal{P}}(x, x', m + k)$ satisfies $T^{\infty}[p |x| ... p |x| + 7\tau) = T^{\infty}[\operatorname{SA}[b + k] ... \operatorname{SA}[b + k] + 7\tau)$. Consider any $k \in [1 ... d]$. First, recall (see above) that $\operatorname{Occ}_{\ell}(P,T) = \{s |x| : s \in \mathsf{S}, \ x \preceq T^{\infty}[s 7\tau ...s) \prec x', \ \text{and} \ y \preceq T^{\infty}[s ... s + 7\tau) \prec yc^{\infty}\}$. This implies that $\operatorname{SA}[b + k] + |x| \in \mathsf{S}$. Denote $j = \operatorname{SA}[b + k] + |x|$ and let $y' = T^{\infty}[j ...j + 7\tau)$. We claim that $m + k \in (\mathsf{weight-count}_{\mathcal{P}}(x, x', y') ... \mathsf{weight-count}_{\mathcal{P}}^{\infty}(x, x', y')]$.

 - We now show $m+k \leq \text{weight-count}_{\widehat{P}}(x,x',y')$. By Lemma 8.23(2), weight-count}_{\widehat{P}}(x,x',y') = $|\{s \in S : x \leq T^{\infty}[s-7\tau \ldots s) \prec x' \text{ and } T^{\infty}[s \ldots s+7\tau) \leq y'\}|$. Thus, similarly as above, weight-count}_{\widehat{P}}(x,x',y') = $|S_{\text{low}}| + |S_{\text{mid}}''| = |S_{\text{low}}| + |P_{\text{mid}}''|$, where $S_{\text{mid}}'' = \{s \in S : x \leq T^{\infty}[s-7\tau \ldots s) \prec x' \text{ and } y \leq T^{\infty}[s \ldots s+7\tau) \leq y'\}$ and $P_{\text{mid}}'' = \{s |x| : s \in S_{\text{mid}}''\}$. Thus, it remains to show $k \leq |P_{\text{mid}}''|$. As before, we have $P_{\text{mid}}'' \subseteq \text{Occ}_{\ell}(P,T)$, except now we have that $SA[b+k'] \in P_{\text{mid}}''$ for at least $k' \in [1 \ldots k]$. Thus, $k \leq |P_{\text{mid}}''|$.

By the above, for every $p \in \mathsf{weight\text{-}select}_{\mathcal{P}}(x,x',m+k)$, it holds $T^{\infty}[p\mathinner{\ldotp\ldotp\ldotp} p+7\tau) = y'$ and $x \preceq \overline{T^{\infty}[p-7\tau\mathinner{\ldotp\ldotp\ldotp} p)} \prec x'$, or equivalently, $T^{\infty}[p-|x|\mathinner{\ldotp\ldotp\ldotp} p+7\tau) = \overline{x}y'$. On the other hand, we have $T^{\infty}[\mathsf{SA}[b+k]+|x|\mathinner{\ldotp\ldotp\ldotp} \mathsf{SA}[b+k]+|x|+7\tau) = y'$ and (by $\mathsf{SA}[b+k] \in \mathsf{Occ}_{\ell}(P,T)$) $T^{\infty}[\mathsf{SA}[b+k]\mathinner{\ldotp\ldotp\ldotp} \mathsf{SA}[b+k]+|x|) = \overline{x}$. Thus, $T^{\infty}[p-|x|\mathinner{\ldotp\ldotp\ldotp} p+7\tau) = T^{\infty}[\mathsf{SA}[b+k]\mathinner{\ldotp\ldotp\ldotp} \mathsf{SA}[b+k]+|x|+7\tau)$.

3. Applying the above step for k=i-b, we obtain that for every $p\in \mathsf{weight\text{-}select}_{\mathcal{P}}(x,x',m+(i-b))$, it holds $T^{\infty}[p-|x|\dots p-|x|+7\tau)=T^{\infty}[\mathsf{SA}[b+k]\dots\mathsf{SA}[b+k]+7\tau)=T^{\infty}[\mathsf{SA}[i]\dots\mathsf{SA}[i]+7\tau)$. By $2\ell\leq 7\tau$, we thus obtain $T^{\infty}[p-|x|\dots p-|x|+2\ell)=T^{\infty}[\mathsf{SA}[i]\dots\mathsf{SA}[i]+2\ell)$, which by Lemma 8.10 and $\mathsf{SA}[i]\in \mathsf{Occ}_{2\ell}(P,T)$ is equivalent to $p-|x|\in \mathsf{Occ}_{2\ell}(P,T)$.

Lemma 8.36. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let S be a τ -synchronizing set of T and $S_{\text{comp}} = \text{comp}_{14\tau}(S, T)$. Let $i \in [1..n]$ be such that $SA[i] \in [1..n-3\tau+2] \setminus R(\tau,T)$. Let $j \in Occ_{\ell}(SA[i],T)$, $c = \max \Sigma$, $s = \text{succ}_{S}(j)$, and x, x', y be such that:

- $\bullet \ \overline{x} = T[j \dots s),$
- $\bullet \ \ x'=xc^{\infty},$
- $\overline{x}y = T[j ... \min(n+1, j+\ell)).$

Let us also denote $\mathcal{P} = \operatorname{StrStrPoints}_{7\tau}(S_{\operatorname{comp}}, T)$ (Definition 6.3), $m = \operatorname{weight-count}_{\mathcal{P}}(x, x', y)$, and $b = \operatorname{RangeBeg}_{\ell}(\operatorname{SA}[i], T)$. Then:

- It holds $m + (i b) \in [1 ... \text{ weight-count}_{\mathcal{P}}(x, x')].$
- Every $p \in \text{weight-select}_{\mathcal{P}}(x, x', m + (i b))$ satisfies $p |x| \in \text{Occ}_{2\ell}(SA[i], T)$.

Proof. Denote $P = T[\operatorname{SA}[i] \dots n]$ and $m = |P| = n - \operatorname{SA}[i] + 1$. Note that $\operatorname{SA}[i] \in [1 \dots n - 3\tau + 2]$ and $\operatorname{SA}[i] \notin \mathbb{R}(\tau,T)$ imply that $m \geq 3\tau - 1$ and $\operatorname{per}(P[1 \dots 3\tau - 1]) > \frac{1}{3}\tau$. Thus, P is τ -nonperiodic. Next, observe that since T[n] does not occur in $T[1 \dots n)$, no nonempty suffix of T is a proper prefix of P. Denote $D = T[j \dots s + 2\tau)$. By Lemma 8.22(1), $3\tau - 1 \leq \ell$, and $j \in \operatorname{Occ}_{\ell}(\operatorname{SA}[i],T) = \operatorname{Occ}_{\ell}(P,T)$, it follows that $j \in [1 \dots n - 3\tau + 2] \setminus \mathbb{R}(\tau,T)$ and $T[j \dots \operatorname{succ}_{\mathbb{S}}(j) + 2\tau) = T[j \dots s + 2\tau) = D$ is a prefix of P. Note that $j \notin \mathbb{R}(\tau,T)$ implies that $D \in \mathcal{D}(\tau,T,\mathbb{S})$. Next, note that the following properties hold for x and y.

- First, we observe that by $\overline{x} = T[j ... s)$, the string \overline{x} is a prefix of D. Note also that by $|D| = s + 2\tau j$ and $|\overline{x}| = s j$, it holds $|\overline{x}| + 2\tau = |D|$.
- Second, we prove that it holds $\overline{x}y = P[1 ... \min(m, \ell)]$. Consider two cases:
 - First, assume j = SA[i]. By $\min(n+1, j+\ell) j = \min(n+1, SA[i] + \ell) SA[i] = \min(n+1 SA[i], \ell) = \min(m, \ell)$ we then obtain $\overline{x}y = T[j ... \min(n+1, j+\ell)) = T[SA[i] ... \min(n+1, SA[i] + \ell)) = P[1 ... \min(n+1, SA[i] + \ell) SA[i]] = P[1 ... \min(m, \ell)].$
 - Let us now thus assume $j \neq \mathrm{SA}[i]$. The assumption $j \in \mathrm{Occ}_{\ell}(\mathrm{SA}[i], T)$ and Lemma 8.11 then imply that $\mathrm{LCE}_T(j, \mathrm{SA}[i]) \geq \ell$, i.e., $T[j \ldots j + \ell) = T[\mathrm{SA}[i] \ldots \mathrm{SA}[i] + \ell)$. This in turn implies $\min(n+1, j+\ell) j \geq \ell$, $m \geq \ell$, and $\overline{x}y = T[j \ldots \min(n+1, j+\ell)) = T[j \ldots j+\ell) = T[\mathrm{SA}[i] \ldots \mathrm{SA}[i] + \ell) = P[1 \ldots \ell] = P[1 \ldots \ell]$.

By definition, it follows that $b = \text{RangeBeg}_{\ell}(\text{SA}[i], T) = \text{RangeBeg}_{\ell}(T[\text{SA}[i] ... n], T) = \text{RangeBeg}_{\ell}(P, T)$ and $\text{SA}[i] \in \text{Occ}_{2\ell}(\text{SA}[i], T) = \text{Occ}_{2\ell}(T[\text{SA}[i] ... n], T) = \text{Occ}_{2\ell}(P, T)$. It thus follows by Lemma 8.35 that:

- It holds $m + (i b) \in [1 ... \text{ weight-count}_{\mathcal{P}}(x, x')],$
- Every $p \in \text{weight-select}_{\mathcal{P}}(x, x', m + (i b))$ satisfies $p |x| \in \text{Occ}_{2\ell}(P, T) = \text{Occ}_{2\ell}(\text{SA}[i], T)$.

Query Algorithms

Proposition 8.37. Let $k \in [4..\lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let $i \in [1..n]$ be such that $SA[i] \in [1..n] \setminus R(\tau, T)$. Given CompSANonperiodic(T) along with the values k, i, RangeBeg $_{\ell}(SA[i], T)$, RangeEnd $_{\ell}(SA[i], T)$, and some position $j \in Occ_{\ell}(SA[i], T)$ satisfying $j = \min Occ_{2\ell}(j, T)$ as input, we can compute some $j' \in Occ_{2\ell}(SA[i], T)$ in $O(\log^{3+\epsilon} n)$ time.

Proof. Denote $b = \text{RangeBeg}_{\ell}(SA[i], T)$ and $e = \text{RangeEnd}_{\ell}(SA[i], T)$, If $e - b = |\text{Occ}_{\ell}(SA[i], T)| = 1$ then by $\text{Occ}_{2\ell}(SA[i], T) \subseteq \text{Occ}_{\ell}(SA[i], T)$, it holds $\text{Occ}_{2\ell}(SA[i], T) = \{j\}$, and hence we return j' = j.

Let us thus assume e-b>1. By the uniqueness of T[n] in T, this implies that $T^{\infty}[j\mathinner{.\,.} j+\ell)$ occurs in T^{∞} starting in at least two different positions in $[1\mathinner{.\,.} n]$, and hence it does not contain T[n]. Thus, $\mathrm{SA}[i]\in[1\mathinner{.\,.} n-\ell]$. By $3\tau-1\le\ell$, we thus have $\mathrm{SA}[i]\in[1\mathinner{.\,.} n-3\tau+2]\setminus \mathrm{R}(\tau,T)$. Note that by $j\in\mathrm{Occ}_{\ell}(\mathrm{SA}[i],T)\subseteq\mathrm{Occ}_{3\tau-1}(\mathrm{SA}[i],T)$ and Lemma 8.22(2), we then obtain $j\in[1\mathinner{.\,.} n-3\tau+2]\setminus \mathrm{R}(\tau,T)$. Observe that using $\mathrm{CompSANonperiodic}(T)$, we can perform LCE_T and $\mathrm{LCE}_{\overline{T}}$ queries in $\mathcal{O}(\log n)$ time, and we can access any symbol of T in $\mathcal{O}(\log n)$ time. Thus, we can compare any two substrings of T^{∞} or T^{∞} (specified with their starting positions and lengths) in $t_{\mathrm{cmp}}=\mathcal{O}(\log n)$ time. The algorithm consists of three steps:

- 1. First, using Proposition 8.26, we compute $s = \operatorname{succ}_{S_k}(j)$ in $\mathcal{O}(\log n)$ time (note that we utilize $j \in \operatorname{Occ}_{\ell}(j,T)$).
- 2. Denote $\mathcal{P} = \operatorname{StrStrPoints}_{7\tau}(\mathsf{S}_{\operatorname{comp},k},T)$ and let x,x',y be such that $\overline{x} = T[j..s), \ x' = xc^{\infty}$, and $\overline{x}y = T[j..\min(n+1,j+\ell))$ (where $c = \max \Sigma$). Using Proposition 6.4(1), in $\mathcal{O}(\log^{2+\epsilon} n + t_{\operatorname{cmp}} \log n) = \mathcal{O}(\log^{2+\epsilon} n)$ time we compute $m := \operatorname{weight-count}_{\mathcal{P}}(x,x',y)$ using parameters $(i,q_l,q_r) = (s,s-j,\min(n+1,j+\ell)-s)$.
- 3. Finally, using Proposition 6.4(2) with parameters $(i, q_l, r) = (s, s j, m + (i b))$ we compute some $p \in \mathsf{weight\text{-}select}_{\mathcal{P}}(x, x', r)$ in $\mathcal{O}(\log^{3+\epsilon} n + t_{\rm cmp} \log n) = \mathcal{O}(\log^{3+\epsilon} n)$ time. By Lemma 8.36, it holds $p |x| \in \mathrm{Occ}_{2\ell}(\mathrm{SA}[i], T)$. We thus return j' := p |x| = p (s j) as the answer.

In total, we spend $\mathcal{O}(\log^{3+\epsilon} n)$ time.

8.3.8 Computing a Position in a Cover

Combinatorial Properties

Lemma 8.38. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let S be a τ -synchronizing set of T and $S_{\text{comp}} = \text{comp}_{14\tau}(S, T)$. Let $P \in \Sigma^m$ be a τ -nonperiodic pattern such that $m \geq 3\tau - 1$, no nonempty suffix of T is a proper prefix of P, and P is prefixed with some $D \in \mathcal{D}(\tau, T, S)$ (Definition 8.19). Let $d \in [\ell ..2\ell]$, $c = \max \Sigma$, and x, x', y, y' be such that:

- \overline{x} is a prefix of D and $|\overline{x}| + 2\tau = |D|$,
- $x' = xc^{\infty}$,
- $\overline{x}y = P[1 .. \min(m, d)],$
- $y' = yc^{\infty}$.

Then, $\min \operatorname{Occ}_d(P,T) = \operatorname{r-min}_{\mathcal{P}}(x,x',y,y') - |x|$, where $\mathcal{P} = \operatorname{StrStrPoints}_{7\tau}(\mathsf{S}_{\operatorname{comp}},T)$ (Definition 6.3).

Proof. The result follows by putting together Lemmas 8.24 and 8.31, i.e.,

$$\min \operatorname{Occ}_{d}(P,T) = \min\{s - |x| : s \in S, \ x \leq \overline{T^{\infty}[s - 7\tau \dots s)} \prec x', \text{ and } y \leq T^{\infty}[s \dots s + 7\tau) \prec y'\}$$

$$= \min\{s \in S : x \leq \overline{T^{\infty}[s - 7\tau \dots s)} \prec x', \text{ and } y \leq T^{\infty}[s \dots s + 7\tau) \prec y'\} - |x|$$

$$= \operatorname{r-min}_{\mathcal{P}}(x, x', y, y') - |x|.$$

Lemma 8.39. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let S be a τ -synchronizing set of T and $S_{\text{comp}} = \text{comp}_{14\tau}(S, T)$. Let $j \in [1..n - 3\tau + 2] \setminus R(\tau, T)$, $c = \max \Sigma$, $s = \text{succ}_S(j)$, $d \in [\ell .. 2\ell]$, and x, x', y, y' be such that:

- $\overline{x} = T[j \dots s),$
- $x' = xc^{\infty}$,
- $\overline{x}y = T[j .. \min(n+1, j+d)),$
- $y' = yc^{\infty}$.

Then, $\min \operatorname{Occ}_d(j,T) = \operatorname{r-min}_{\mathcal{P}}(x,x',y,y') - |x|$, where $\mathcal{P} = \operatorname{StrStrPoints}_{7\tau}(\mathsf{S}_{\operatorname{comp}},T)$ (Definition 6.3).

Proof. The result follows by Lemma 8.38 (see the proof of Lemma 8.32 for a similar argument). \Box

Query Algorithms

Proposition 8.40. Let $k \in [4..\lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $j \in [1..n] \setminus \mathsf{R}(\tau,T)$. Given the structure CompSANonperiodic(T), the value k, the position j, and any $j' \in \mathrm{Occ}_{\ell}(j,T)$ satisfying $j' = \min \mathrm{Occ}_{2\ell}(j',T)$ as input, we can compute $\min \mathrm{Occ}_{2\ell}(j,T)$ in $\mathcal{O}(\log^2 n)$ time.

Proof. If $j > n - 3\tau + 2$, then we have $|\operatorname{Occ}_{\ell}(j,T)| = 1$. Thus, by $j \in \operatorname{Occ}_{\ell}(j,T)$ and $\operatorname{Occ}_{2\ell}(j,T) \subseteq \operatorname{Occ}_{\ell}(j,T)$ we return that $\min \operatorname{Occ}_{2\ell}(j,T) = j$.

Let us thus assume $j \in [1 ... n - 3\tau + 2]$. The algorithm consists of two steps:

1. Using Proposition 8.26, we compute $s = \operatorname{succ}_{S_k}(j)$ in $\mathcal{O}(\log n)$ time.

2. Let x, x', y, y' be such that $T^{\infty}[j ... s) = \overline{x}$, $x' = xc^{\infty}$, $T^{\infty}[j ... j + 2\ell) = \overline{x}y$, and $y' = yc^{\infty}$ (where $c = \max \Sigma$). By Lemma 8.39, it then holds $\min \operatorname{Occ}_{2\ell}(j,T) = \operatorname{r-min}_{\mathcal{P}}(x,x',y,y') - |x|$, where $\mathcal{P} = \operatorname{StrStrPoints}_{7\tau}(\mathsf{S}_{\operatorname{comp},k},T)$ (Definition 6.3). Observe also that using $\operatorname{CompSANonperiodic}(T)$, we can perform LCE_T and $\operatorname{LCE}_{\overline{T}}$ queries in $\mathcal{O}(\log n)$ time, and we can access any symbol of T in $\mathcal{O}(\log n)$ time. Thus, we can compare any two substrings of T^{∞} or $\overline{T^{\infty}}$ (specified with their starting positions and lengths) in $t_{\operatorname{cmp}} = \mathcal{O}(\log n)$ time. Thus, using the data structure from Proposition 6.4 for $q = 7\tau$ and $P = \mathsf{S}_{\operatorname{comp},k}$ (which is a component of $\operatorname{CompSANonperiodic}(T)$) with parameters $(i, q_l, q_r) = (s, s - j, \min(n+1, j+2\ell) - s)$ we compute $\min \operatorname{Occ}_{2\ell}(j,T)$ in $\mathcal{O}(\log^{1+\epsilon} n + t_{\operatorname{cmp}} \log n) = \mathcal{O}(\log^2 n)$ time. In total, we spend $\mathcal{O}(\log^2 n)$ time.

8.3.9 Implementation of ISA Queries

Proposition 8.41. Let $k \in [4 ... \lceil \log n \rceil)$. Denote $\ell = 2^k$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let $j \in [1 ... n] \setminus R(\tau, T)$. Given CompSANonperiodic(T), the value k, the position j, any $j' \in \operatorname{Occ}_{\ell}(j,T)$ satisfying $j' = \min \operatorname{Occ}_{2\ell}(j',T)$, and (RangeBeg $_{\ell}(j,T)$, RangeEnd $_{\ell}(j,T)$) as input, we can compute (RangeBeg $_{\ell}(j,T)$, RangeEnd $_{\ell}(j,T)$) and some position $j'' \in \operatorname{Occ}_{\ell}(j,T)$ satisfying $j'' = \min \operatorname{Occ}_{\ell}(j'',T)$ in $\mathcal{O}(\log^{2+\epsilon} n)$ time.

Proof. Denote $b = \text{RangeBeg}_{\ell}(j, T)$. The algorithm proceeds in two steps (note that their order can be swapped):

- 1. First, using Propositions 8.30 and 8.34 in $\mathcal{O}(\log^{2+\epsilon} n)$ time we compute $\delta = |\operatorname{Pos}_{\ell}^{\operatorname{beg}}(j,T)|$ and $m = |\operatorname{Occ}_{2\ell}(j,T)|$. We then have $(\operatorname{RangeBeg}_{2\ell}(j,T),\operatorname{RangeEnd}_{2\ell}(j,T)) = (b+\delta,b+\delta+m)$.
- 2. Second, using Proposition 8.40 in $\mathcal{O}(\log^2 n)$ time we compute $j'' = \min \operatorname{Occ}_{2\ell}(j,T)$. Such position clearly satisfies $j'' \in \operatorname{Occ}_{2\ell}(j,T)$ and $j'' = \min \operatorname{Occ}_{4\ell}(j'',T)$.

In total, we spend $\mathcal{O}(\log^{2+\epsilon} n)$ time.

8.3.10 Implementation of SA Queries

Proposition 8.42. Let $k \in [4..\lceil \log n \rceil)$. Denote $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let $i \in [1..n]$ be such that $SA[i] \in [1..n] \setminus R(\tau,T)$. Given CompSANonperiodic(T) along with values k, i, $RangeBeg_{\ell}(SA[i],T)$, $RangeEnd_{\ell}(SA[i],T)$, and some $j \in Occ_{\ell}(SA[i],T)$ satisfying $j = \min Occ_{2\ell}(j,T)$ as input, we can compute the pair $(RangeBeg_{2\ell}(SA[i],T), RangeEnd_{2\ell}(SA[i],T))$ along with some $j' \in Occ_{2\ell}(SA[i],T)$ satisfying $j' = \min Occ_{4\ell}(j',T)$ in $\mathcal{O}(\log^{3+\epsilon} n)$ time.

Proof. Denote $b = \text{RangeBeg}_{\ell}(SA[i], T)$ and $e = \text{RangeEnd}_{\ell}(SA[i], T)$. The algorithm consists of two steps:

- 1. First, using Proposition 8.37 we compute some position $p \in \operatorname{Occ}_{2\ell}(\operatorname{SA}[i],T)$ in $\mathcal{O}(\log^{3+\epsilon}n)$ time. Note that then it holds $\operatorname{Occ}_{\ell}(p,T) = \operatorname{Occ}_{\ell}(\operatorname{SA}[i],T)$ (and hence, in particular, $j \in \operatorname{Occ}_{\ell}(p,T)$), Range $\operatorname{Beg}_{\ell}(p,T) = b$, and Range $\operatorname{End}_{\ell}(p,T) = e$. Moreover, by $3\tau 1 \leq \ell$ and Definition 2.4, we have $p \in [1..n] \setminus \operatorname{R}(\tau,T)$.
- 2. Using Proposition 8.41, we compute $b' = \text{RangeBeg}_{2\ell}(p,T)$, $e' = \text{RangeEnd}_{2\ell}(p,T)$, and some $j' \in \text{Occ}_{2\ell}(p,T)$ satisfying $j' = \min \text{Occ}_{4\ell}(j',T)$ in $\mathcal{O}(\log^{2+\epsilon}n)$ time. By $p \in \text{Occ}_{2\ell}(\text{SA}[i],T)$, we then have $\text{RangeBeg}_{2\ell}(\text{SA}[i],T) = b'$, $\text{RangeEnd}_{2\ell}(\text{SA}[i],T) = e'$, and $j' \in \text{Occ}_{2\ell}(\text{SA}[i],T)$.

In total, we spend $\mathcal{O}(\log^{3+\epsilon} n)$ time.

8.3.11 Construction Algorithm

Proposition 8.43. Given the LZ77 parsing of T, we can construct CompSANonperiodic(T) in $\mathcal{O}(\delta(T)\log^7 n)$ time.

Proof. First, using [KK20b, Theorem 6.11] (resp. [KK20b, Theorem 6.21]), in $\mathcal{O}(z(T)\log^4 n)$ time we construct a structure that, given any substring S of T (specified with its starting position and the length) in $\mathcal{O}(\log^3 n)$ time returns min $\operatorname{Occ}(S,T)$ (resp. $|\operatorname{Occ}(S,T)|$).

We then construct the components of CompSANonperiodic(T) (Section 8.3.2) as follows:

- 1. In $\mathcal{O}(\delta(T)\log^7 n)$ time we construct CompSACore(T) using Proposition 8.18.
- 2. Using Proposition 5.30 with c = 14, in $\mathcal{O}(\delta(T)\log^7 n)$ time we construct $\{S_{\text{comp},k}\}_{k \in [4..\lceil \log n \rceil)}$ (as defined in Section 8.3.2). Then, for $k \in [4.. \lceil \log n \rceil)$, in $\mathcal{O}(n_k \log n_k) = \mathcal{O}(n_k \log n)$ time (where n_k is defined in Section 8.3.2), we sort the elements of $S_{\text{comp},k}$ and write to $A_{\text{comp},k}$. By Proposition 5.30, we have $\sum_{k \in [4..\lceil \log n \rceil)} n_k = \mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$. Thus, over all k, the sorting takes $\mathcal{O}(\delta(T) \log^2 n)$ time. 3. For $k = 4, \ldots, \lceil \log n \rceil - 1$, we construct the structure from Proposition 6.4 with $P = \mathsf{S}_{\mathrm{comp},k}$ and
- $q = 7\tau_k$ as input. Over all k, this takes $\mathcal{O}(\sum_{k \in [4..\lceil \log n \rceil)} n_k \log^3 n) = \mathcal{O}(\delta(T) \log^4 n)$ time (see above). Recall that using CompSACore(T) we can perform LCE_T and LCE_{\overline{T}} queries in $\mathcal{O}(\log n)$ time, and we can access any symbol of T in $\mathcal{O}(\log n)$ time. Thus, we can compare any two substrings of T^{∞} or $\overline{T^{\infty}}$ (specified with their starting positions and lengths) in $\mathcal{O}(\log n)$ time. Consequently, all structures needed in Proposition 6.4 were either constructed above or are part of CompSACore(T).

In total, the construction takes $\mathcal{O}(\delta(T)\log^7 n + z(T)\log^4 n) = \mathcal{O}(\delta(T)\log^7 n)$ time.

The Periodic Patterns and Positions 8.4

Preliminaries 8.4.1

Notation and Definitions for Patterns

Definition 8.44 (Necklace-consistent function). A function $f: \Sigma^+ \to \Sigma^+$ is said to be necklace-consistent if it satisfies the following conditions for every $S, S' \in \Sigma^+$:

- 1. The strings f(S) and S are cyclically equivalent.
- 2. If S and S' are cyclically equivalent, then f(S) = f(S').

Let f be some necklace-consistent function. Let also $\tau \geq 1$ and $P \in \Sigma^m$ be a τ -periodic pattern. Denote $p = \text{per}(P[1..3\tau - 1])$. We define $\text{root}_f(\tau, P) := f(P[1..p])$ and $e(\tau, P) := 1 + p + \text{lcp}(P[1..m], P[1 + p..m])$. Observe that then we can write $P[1..e(\tau,P)] = H'H^kH''$, where $H = \text{root}_f(\tau,P)$, and H' (resp. H'') is a proper suffix (resp. prefix) of H. Note that there is always only one way to write $P[1..e(\tau, P))$ in this way, since the opposite would contradict the synchronization property of primitive strings [CHL07, Lemma 1.11]. We denote $\operatorname{head}_f(\tau, P) := |H'|$, $\exp_f(\tau, P) := k$, and $\operatorname{tail}_f(\tau, P) := |H''|$. For any $t \geq 3\tau - 1$, we define $\exp_f^{\mathrm{cut}}(\tau,P,t) := \min(\exp_f(\tau,P), \lfloor \frac{t-s}{|H|} \rfloor) \text{ and } e_f^{\mathrm{cut}}(\tau,P,t) := 1+s+\exp_f^{\mathrm{cut}}(\tau,P,t) \cdot |H|, \text{ where } s = \mathrm{head}_f(\tau,P)$ and $H = \mathrm{root}_f(\tau,P)$. We denote $e_f^{\mathrm{cut}}(\tau,P) := e_f^{\mathrm{cut}}(\tau,P,m)$. Note that $e_f^{\mathrm{full}}(\tau,P) = j+s+\exp_f(\tau,P) \cdot |H| = e(\tau,P) - \mathrm{tail}_f(\tau,P)$. Next, letting $p = |\mathrm{per}(P[1..3\tau-1])|$, we define $\mathrm{type}(\tau,P) = +1$ if $e(\tau,P) \leq |P|$ and $P[e(\tau, P)] > P[e(\tau, P) - p]$, and type $(\tau, P) = -1$ otherwise.

Lemma 8.45. Let $\tau \geq 1$, $P \in \Sigma^+$ be a τ -periodic pattern, and f be a necklace-consistent function. For every $P' \in \Sigma^+$, $lcp(P, P') > 3\tau - 1$ holds if and only if P' is τ -periodic, $root_f(\tau, P') = root_f(\tau, P)$, and $\operatorname{head}_f(\tau, P') = \operatorname{head}_f(\tau, P)$. Moreover, if $e(\tau, P) \leq |P|$ and $\operatorname{lcp}(P, P') \geq e(\tau, P)$ (which holds, in particular, when P is a prefix of P'), then:

- $e(\tau, P') = e(\tau, P)$,
- $tail_f(\tau, P') = tail_f(\tau, P),$ $e_f^{\text{full}}(\tau, P') = e_f^{\text{full}}(\tau, P),$ $\exp_f(\tau, P') = \exp_f(\tau, P),$
- $\operatorname{type}(\tau, P') = \operatorname{type}(\tau, P)$.

Proof. Denote $s = \text{head}_f(\tau, P)$, $H = \text{root}_f(\tau, P)$, and p = |H|. Let us first assume $lcp(P, P') \geq 3\tau - 1$. This implies $|P'| \ge \text{lcp}(P, P') \ge 3\tau - 1$ and $P'[1...3\tau - 1] = P[1...3\tau - 1]$. Thus, $\text{per}(P'[1...3\tau - 1]) = P[1...3\tau - 1]$ per $(P[1..3\tau-1]) = p \le \frac{1}{3}\tau$, i.e., P' is τ -periodic. Moreover, root_f $(\tau, P') = f(P'[1..p]) = f(P[1..p]) = H$. To show head_f $(\tau, P') = s$, note that by $|H| \le \tau$, the string $H'H^2$ (where H' is a length-s suffix of H) is a prefix of $P[1...3\tau-1] = P'[1...3\tau-1]$. On the other hand, letting $s' = \text{head}_f(\tau, P')$, it holds that $\widehat{H}H^2$ (where H is a length-s' suffix of H) is a prefix of $P'[1..3\tau - 1]$. Thus, by the synchronization property of primitive strings [CHL07, Lemma 1.11] applied to the two copies of H, we have s' = s, i.e., head $f(\tau, P') = \text{head}_f(\tau, P)$. For the converse implication, assume that P' is τ -periodic and it holds $\operatorname{root}_f(\tau, P') = H$ and $\operatorname{head}_f(\tau, P') = s$.

This implies that both $P[1...e(\tau, P))$ and $P'[1...e(\tau, P'))$ are prefixes of $H' \cdot H^{\infty}[1...)$ (where H' is as above). Thus, by $e(\tau, P) - 1$, $e(\tau, P') - 1 \ge 3\tau - 1$, we obtain $lcp(P, P') \ge 3\tau - 1$.

Let us now assume $e(\tau,P) \leq |P|$ and $\operatorname{lcp}(P,P') \geq e(\tau,P)$. Note that $e(\tau,P) > 3\tau - 1$ and the first part of the claim then imply that P' is τ -periodic and it holds $\operatorname{root}_f(\tau,P') = H$ and $\operatorname{head}_f(\tau,P') = s$. By $e(\tau,P) \leq |P|$, it follows that $P[e(\tau,P)] \neq P[e(\tau,P)-p]$. Combining with $\operatorname{lcp}(P,P') \geq e(\tau,P)$, we thus have $\operatorname{lcp}(P[1\mathinner{\ldotp\ldotp}|P|],P[1+p\mathinner{\ldotp\ldotp}|P|]) = \operatorname{lcp}(P'[1\mathinner{\ldotp\ldotp}|P'|],P'[1+p\mathinner{\ldotp\ldotp}|P'|])$. Consequently, $e(\tau,P') = 1+p+\operatorname{lcp}(P'[1\mathinner{\ldotp\ldotp}|P'|],P'[1+p\mathinner{\ldotp\ldotp}|P'|]) = e(\tau,P)$. We then also obtain $\operatorname{tail}_f(\tau,P') = (e(\tau,P')-1-\operatorname{head}_f(\tau,P')) \bmod p = (e(\tau,P)-1-\operatorname{head}_f(\tau,P)) \bmod p_{\text{full}} \cot p = (\tau,P) + \operatorname{tail}_f(\tau,P') = e(\tau,P') - \operatorname{tail}_f(\tau,P') = e(\tau,P) - \operatorname{tail}_f(\tau,P) = e^{\operatorname{full}}(\tau,P) + \operatorname{erp}_f(\tau,P') = \operatorname{lcp}_f(\tau,P') = \operatorname{lc$

Lemma 8.46. Let $\tau \geq 1$, f be a necklace-consistent function, and $P_1, P_2 \in \Sigma^+$ be τ -periodic patterns such that $\operatorname{root}_f(\tau, P_1) = \operatorname{root}_f(\tau, P_2)$ and $\operatorname{head}_f(\tau, P_1) = \operatorname{head}_f(\tau, P_2)$. Denote $t_1 = e(\tau, P_1) - 1$ and $t_2 = e(\tau, P_2) - 1$. Then, it holds $\operatorname{lcp}(P_1, P_2) \geq \min(t_1, t_2)$. Moreover:

- 1. If $type(\tau, P_1) \neq type(\tau, P_2)$ or $t_1 \neq t_2$, then $P_1 \neq P_2$ and $lep(P_1, P_2) = min(t_1, t_2)$,
- 2. If $\operatorname{type}(\tau, P_1) \neq \operatorname{type}(\tau, P_2)$, then $P_1 \prec P_2$ if and only if $\operatorname{type}(\tau, P_1) < \operatorname{type}(\tau, P_2)$,
- 3. If $type(\tau, P_1) = -1$, then $t_1 < t_2$ implies $P_1 \prec P_2$,
- 4. If $type(\tau, P_1) = +1$, then $t_1 < t_2$ implies $P_1 > P_2$,
- 5. If $\operatorname{type}(\tau, P_1) = \operatorname{type}(\tau, P_2) = -1$ and $t_1 \neq t_2$, then $t_1 < t_2$ if and only if $P_1 \prec P_2$,
- 6. If $\operatorname{type}(\tau, P_1) = \operatorname{type}(\tau, P_2) = +1$ and $t_1 \neq t_2$, then $t_1 < t_2$ if and only if $P_1 > P_2$.

Proof. Denote $H = \operatorname{root}_f(\tau, P_1) = \operatorname{root}_f(\tau, P_2)$, $s = \operatorname{head}_f(\tau, P_1) = \operatorname{head}_f(\tau, P_2)$, and p = |H|. Let H' be a length-s suffix of H and let $Q = H'H^{\infty}$. We first observe that both $P_1[1 \dots e(\tau, P_1))$ and $P_2[1 \dots e(\tau, P_2))$ are prefixes of Q, i.e., $P_1[1 \dots e(\tau, P_1)) = P_1[1 \dots t_1] = Q[1 \dots t_1]$ and $P_2[1 \dots e(\tau, P_2)) = P_2[1 \dots t_2] = Q[1 \dots t_2]$. Thus, $\operatorname{lcp}(P_1, P_2) \geq \min(t_1, t_2)$.

We now prove the remaining claims:

- 1. Recall that $P_1[1 ... e(\tau, P_1)) = P_1[1 ... t_1] = Q[1 ... t_1]$, $P_2[1 ... e(\tau, P_2)) = P_2[1 ... t_2] = Q[1 ... t_2]$, and $lcp(P_1, P_2) \ge min(t_1, t_2)$. Thus, it remains to prove $P_1 \ne P_2$ and $lcp(P_1, P_2) \le min(t_1, t_2)$. Let us first assume $t_1 \ne t_2$. Without the loss of generality, let $t_1 < t_2$. Consider two cases:
 - First, assume $e(\tau, P_1) = |P_1| + 1$. Then, $P_1 = Q[1 ... t_1]$. This implies $lcp(P_1, P_2) \le t_1 = min(t_1, t_2)$. On the other hand, from $|P_2| \ge t_2 > t_1 = |P_1|$, we then obtain $P_1 \ne P_2$.
 - Let us now assume $e(\tau, P_1) \leq |P_1|$. By $t_1 < t_2$, this implies $P_1[1+t_1] = P_1[e(\tau, P_1)] \neq P_1[e(\tau, P_1) p] = P_1[1+t_1-p] = Q[1+t_1-p] = Q[1+t_1] = P_2[1+t_1]$. Thus, $P_1 \neq P_2$ and $lcp(P_1, P_2) \leq t_1 = min(t_1, t_2)$.

Let us now assume $t_1 = t_2$ and $\operatorname{type}(\tau, P_1) \neq \operatorname{type}(\tau, P_2)$. Without the loss of generality, let us assume $\operatorname{type}(\tau, P_1) = -1$ and $\operatorname{type}(\tau, P_2) = +1$. Note that then $e(\tau, P_2) \leq |P_2|$ and $P_2[e(\tau, P_2)] \succ P_2[e(\tau, P_2) - p]$. Consider two cases:

- First, assume $e(\tau, P_1) = |P_1| + 1$. Then, it holds $P_1 = Q[1..t_1]$. This immediately implies $lcp(P_1, P_2) \le t_1 = \min(t_1, t_2)$. On the other hand, $type(\tau, P_2) = +1$ implies $|P_2| \ge e(\tau, P_2) = t_2 + 1 > t_1 = |P_1|$. Thus, $P_1 \ne P_2$.
- Let us now assume $e(\tau, P_1) \leq |P_1|$. By $\operatorname{type}(\tau, P_1) = -1$, we then must have $P_1[e(\tau, P_1)] \prec P_1[e(\tau, P_1) p]$. Recall that in the proof of Lemma 8.46(1), we observed that in this situation it holds $P_1[1 + t_1] \prec Q[1 + t_1]$ and $Q[1 + t_2] \prec P_2[1 + t_2]$. By $t_1 = t_2$, this implies $P_1 \neq P_2$ and $\operatorname{lcp}(P_1, P_2) \leq \min(t_1, t_2)$.
- 2. Let us first assume type $(\tau, P_1) = -1$ and type $(\tau, P_2) = +1$. We will prove that $P_1 \prec Q \prec P_2$. Recall that $P_1[1 \ldots e(\tau, P_1)) = P_1[1 \ldots t_1] = Q[1 \ldots t_1]$. If $t_1 = |P_1|$ then we immediately obtain $P_1 \prec Q$. Otherwise, type $(\tau, P_1) = -1$ implies that $t_1 < |P_1|$ and $P_1[1 + t_1] \prec P_1[1 + t_1 p] = Q[1 + t_1 p] = Q[1 + t_1]$. Thus, we again have $P_1 \prec Q$. Next, recall that $P_2[1 \ldots e(\tau, P_2)) = P_2[1 \ldots t_2] = Q[1 \ldots t_2]$. On the other hand, type $(\tau, P_2) = +1$ implies that $e(\tau, P_2) \leq |P_2|$ and $P_2[1 + t_2] \succ P_2[1 + t_2 p] = Q[1 + t_2 p] = Q[1 + t_2]$.

- Hence, $P_2 \succ Q$. We have thus proved $P_1 \prec Q \prec P_2$, which implies $P_1 \prec P_2$. Let us now assume $P_1 \prec P_2$. Suppose $\operatorname{type}(\tau, P_1) \geq \operatorname{type}(\tau, P_2)$. By the assumption $\operatorname{type}(\tau, P_1) \neq \operatorname{type}(\tau, P_2)$, we then must have $\operatorname{type}(\tau, P_1) = +1$ and $\operatorname{type}(\tau, P_2) = -1$. Analogously as above, it then follows $P_2 \prec Q \prec P_1$, which contradicts the assumption. Thus, type $(\tau, P_1) < \text{type}(\tau, P_2)$.
- 3. Recall that $P_1[1...e(\tau, P_1)) = P_1[1...t_1] = Q[1...t_1]$ and $P_2[1...e(\tau, P_2)) = P_2[1...t_2] = Q[1...t_2]$. If $e(\tau, P_1) = |P_1| + 1$, then $P_1 = P_1[1 ... e(\tau, P_1)) = Q[1 ... t_1]$. By $t_1 < t_2$, we thus obtain $P_1 = P_1[1 ... e(\tau, P_1)]$ $Q[1..t_1] \prec Q[1..t_2] = P_2[1..e(\tau, P_2)) \leq P_2$. Let us thus assume $e(\tau, P_1) \leq |P_1|$. By $t_1 < t_2$, we have $P_1[1..t_1] = P_2[1..t_1]$. On the other hand, by $type(\tau, P_1) = -1$ and $t_1 < t_2$, we have $P_1[1+t_1] = P_1[e(\tau, P_1)] \prec P_1[e(\tau, P_1) - p] = P_1[1+t_1-p] = Q[1+t_1-p] = Q[1+t_1] = P_2[1+t_1].$ We thus obtain $P_1[1..1+t_1] \prec P_2[1..1+t_1]$, which implies $P_1 \prec P_2$.
- 4. Recall (as above) that $P_1[1 ... e(\tau, P_1)) = P_1[1 ... t_1] = Q[1 ... t_1]$ and $P_2[1 ... e(\tau, P_2)) = P_2[1 ... t_2] = P_2[1 ... t_2]$ $Q[1...t_2]$. The assumptions type $(\tau, P_1) = +1$ and $t_1 < t_2$ imply $P_1[1+t_1] = P_1[e(\tau, P_1)] > P_1[e(\tau, P_1)] - P_1[e(\tau, P_1)] > P_1[$ $[p] = P_1[1+t_1-p] = Q[1+t_1-p] = Q[1+t_1] = P_2[1+t_1]$. We thus obtain $P_1[1..1+t_1] > P_2[1..1+t_1]$, which implies $P_1 \succ P_2$.
- 5. The first implication follows by Lemma 8.46(3). Let us thus assume $P_1 \prec P_2$. Suppose $t_1 \geq t_2$. The assumption $t_1 \neq t_2$ then implies $t_1 > t_2$. Recall that it holds $P_1[1 \dots e(\tau, P_1)) = P_1[1 \dots t_1] = Q[1 \dots t_1]$ and $P_2[1...e(\tau, P_2)) = P_2[1...t_2] = Q[1...t_2]$. By $t_1 > t_2$, we thus have $P_1[1...t_2] = P_2[1...t_2]$. Consider two cases. If $e(\tau, P_2) + 1 = |P_2|$, then $P_2 = P_2[1 ... e(\tau, P_2)) = Q[1 ... t_2] \prec Q[1 ... t_1] = P_1[1 ... e(\tau, P_1)) \leq P_1$, a contradiction. Assume now $e(\tau, P_2) \leq |P_2|$. Then, $P_2[1 + t_2] = P_2[e(\tau, P_2)] \prec P_2[e(\tau, P_2) - p] =$ $P_2[1+t_2-p] = Q[1+t_2-p] = Q[1+t_2] = P_1[1+t_2]. \text{ Therefore, we obtain } P_2[1\dots 1+t_2] \prec P_1[1\dots 1+t_2],$ which again implies $P_2 \prec P_1$, a contradiction. Thus, we must have $t_1 < t_2$.
- 6. The first implication follows by Lemma 8.46(4). Let us thus assume $P_1 > P_2$. Suppose $t_1 \ge t_2$. The assumption $t_1 \neq t_2$ then implies $t_1 > t_2$. The assumptions type $(\tau, P_2) = +1$ and $t_1 > t_2$ imply $e(\tau, P_2) \leq$ $|P_2|$ and $P_2[1+t_2] = P_2[e(\tau, P_2)] > P_2[e(\tau, P_2) - p] = P_2[1+t_2-p] = Q[1+t_2-p] = Q[1+t_2] = P_1[1+t_2]$. We thus obtain $P_2[1..1+t_2] > P_1[1..1+t_2]$, which implies $P_2 > P_1$, contradicting the assumption. Thus, we must have $t_1 < t_2$.

Lemma 8.47. Let $\tau \geq 1$, $P \in \Sigma^+$ be a τ -periodic pattern, and f be a necklace-consistent function. Let $t \geq t' \geq 3\tau - 1$ and $P' = P[1 ... \min(|P|, t)]$. Then, P' is τ -periodic and it holds:

```
1. \exp_f^{\text{cut}}(\tau, P, t') = \exp_f^{\text{cut}}(\tau, P', t'),
```

- 2. $\exp_f^{\text{cut}}(\tau, P, t) = \exp_f(\tau, P'),$ 3. $e_f^{\text{cut}}(\tau, P, t') = e_f^{\text{cut}}(\tau, P', t'),$ 4. $e_f^{\text{cut}}(\tau, P, t) = e_f^{\text{full}}(\tau, P').$

Proof. 1. Denote m = |P|, $s = \text{head}_f(\tau, P)$, $H = \text{root}_f(\tau, P)$, p = |H|, and $k_1 = \exp_f^{\text{cut}}(\tau, P, t') = \exp_f^{\text{cut}}(\tau, P, t')$ $\min(\exp_f(\tau, P), \lfloor \frac{t'-s}{n} \rfloor)$. Observe that by $m \geq 3\tau - 1$ and $t \geq 3\tau - 1$, we have $|P'| = \min(m, t) \geq 3\tau - 1$. Thus, by Lemma 8.45, P' is τ -periodic and it holds head $f(\tau, P') = s$ and root $f(\tau, P') = H$. We first show $\exp_f^{\text{cut}}(\tau, P', t') = k_1$. Consider two cases:

- First, assume $t \leq e(\tau, P) 1$. This implies $\exp_f(\tau, P) = \lfloor \frac{e(\tau, P) 1 s}{p} \rfloor \geq \lfloor \frac{t s}{p} \rfloor \geq \lfloor \frac{t' s}{p} \rfloor$. Thus, $k_1 = \min(\exp_f(\tau, P), \lfloor \frac{t' s}{p} \rfloor) = \lfloor \frac{t' s}{p} \rfloor$. On the other hand, we then have $m \geq e(\tau, P) 1 \geq t$ and hence $|P'| = \min(m, t) = t$. Observe that by $p + \text{lcp}(P[1 ...m], P[1 + p ...m]) = e(\tau, P) - 1 \ge t$, we then also have $p + \text{lcp}(P[1 \dots t], P[1+p \dots t]) = t$. Thus, $e(\tau, P') - 1 = p + \text{lcp}(P'[1 \dots t], P'[1+p \dots t]) = p + \text{lcp}(P[1 \dots t], P[1+p \dots t]) = t$. This in turn implies $\exp_f(\tau, P') = \lfloor \frac{e(\tau, P') - 1 - s}{p} \rfloor = \lfloor \frac{t - s}{p} \rfloor \geq \lfloor \frac{t' - s}{p} \rfloor$. Consequently, $\exp_f^{\text{cut}}(\tau, P', t') = \min(\exp_f(\tau, P'), \lfloor \frac{t' - s}{p} \rfloor) = \lfloor \frac{t' - s}{p} \rfloor = k_1$.

 • Let us now assume $e(\tau, P) - 1 < t$. Consider two subcases. If m < t, then P' = P and we
- immediately obtain $\exp_f^{\text{cut}}(\tau, P', t') = \exp_f^{\text{cut}}(\tau, P, t')$. Otherwise (i.e., $m \ge t$), we have $e(\tau, P) \le m$ and $lcp(P', P) = t \ge e(\tau, P)$, and thus by Lemma 8.45, it holds $exp_f(\tau, P') = exp_f(\tau, P)$. This implies $\exp_f^{\text{cut}}(\tau, P', t') = \exp_f^{\text{cut}}(\tau, P, t').$

We prove the remaining claims as follows:

• To prove the second claim, observe that by $t \ge |P'| \ge e(\tau, P') - 1$ it follows that $\exp_f^{\text{cut}}(\tau, P', t) = \min(\exp_f(\tau, P'), \lfloor \frac{t-s}{p} \rfloor) = \min(\lfloor \frac{e(\tau, P') - 1 - s}{p} \rfloor, \lfloor \frac{t-s}{p} \rfloor) = \lfloor \frac{e(\tau, P') - 1 - s}{p} \rfloor = \exp_f(\tau, P')$. Thus, we obtain by

 $\begin{array}{l} \operatorname{Lemma~8.47(1)~that~exp_f^{\mathrm{cut}}(\tau,P,t) = exp_f^{\mathrm{cut}}(\tau,P',t) = exp_f(\tau,P').} \\ \bullet \ \operatorname{By~Lemma~8.47(1)}, \ e_f^{\mathrm{cut}}(\tau,P,t') = s + \exp_f^{\mathrm{cut}}(\tau,P,t')p = s + \exp_f^{\mathrm{cut}}(\tau,P',t')p = e_f^{\mathrm{cut}}(\tau,P',t').} \\ \bullet \ \operatorname{By~Lemma~8.47(2)}, \ e_f^{\mathrm{cut}}(\tau,P,t) = s + \exp_f^{\mathrm{cut}}(\tau,P,t)p = s + \exp_f(\tau,P')p = e_f^{\mathrm{full}}(\tau,P'). \end{array}$

Notation and Definitions for Positions Let $\tau \in [1... | \frac{n}{2}]$ and f be some necklace-consistent function. Observe that if $j \in R(\tau, T)$, then T[j ... n] is τ -periodic (Definition 8.1). Letting P = T[j ... n]and $p = per(T[j...j + 3\tau - 1))$, we denote $root_f(\tau, T, j) := root_f(\tau, P)$, $e(\tau, T, j) := j + e(\tau, P) - 1 =$ $j+p+\mathrm{LCE}_T(j,j+p),\ \mathrm{head}_f(\tau,T,j):=\mathrm{head}_f(\tau,P),\ \mathrm{exp}_f(\tau,T,j):=\exp_f(\tau,P),\ \mathrm{tail}_f(\tau,T,j):=\mathrm{tail}_f(\tau,P),\\ \exp_f^{\mathrm{cut}}(\tau,T,j,t):=\exp_f^{\mathrm{cut}}(\tau,P,t),\ e_f^{\mathrm{cut}}(\tau,T,j,t):=j+e_f^{\mathrm{cut}}(\tau,P,t)-1\ (\mathrm{with}\ t\geq 3\tau-1),\ e_f^{\mathrm{full}}(\tau,T,j):=j+e_f^{\mathrm{full}}(\tau,P)-1,\ \mathrm{and}\ \mathrm{type}(\tau,T,j):=\mathrm{type}(\tau,P).$ Note that $e_f^{\mathrm{full}}(\tau,T,j)=j+s+\exp_f(\tau,T,j)|H|=e(\tau,T,j)-\mathrm{tail}_f(\tau,T,j),\ \mathrm{where}\ s=\mathrm{head}_f(\tau,T,j)\ \mathrm{and}\ H=\mathrm{root}_f(\tau,T,j).$ We also let $b(\tau,T,j)=j-1$ lcs(T[1..j), T[1..j+p)).

Let $H \in \Sigma^+$ and $s \in \mathbb{Z}_{>0}$. We will repeatedly refer to the following subsets of $R(\tau, T)$:

- $R^-(\tau, T) := \{ j \in R(\tau, T) : type(\tau, T, j) = -1 \},$
- $R^+(\tau,T) := R(\tau,T) \setminus R^-(\tau,T)$,
- $\mathsf{R}_{f,H}(\tau,T) := \{ j \in \mathsf{R}(\tau,T) : \mathsf{root}_f(\tau,T,j) = H \},$
- $\bullet \ \mathsf{R}^-_{f,H}(\tau,T) := \mathsf{R}^-(\tau,T) \cap \mathsf{R}_{f,H}(\tau,T),$
- $\mathsf{R}^{+}_{f,H}(\tau,T) := \mathsf{R}^{+}(\tau,T) \cap \mathsf{R}_{f,H}(\tau,T),$
- $\bullet \ \mathsf{R}_{f,s,H}(\tau,T) := \{j \in \mathsf{R}_{f,H}(\tau,T) : \mathrm{head}_f(\tau,T,j) = s\},$
- $\bullet \ \mathsf{R}^-_{f,s,H}(\tau,T) := \mathsf{R}^-(\tau,T) \cap \mathsf{R}_{f,s,H}(\tau,T), \\ \bullet \ \mathsf{R}^+_{f,s,H}(\tau,T) := \mathsf{R}^+(\tau,T) \cap \mathsf{R}_{f,s,H}(\tau,T),$
- $\mathsf{R}_{f,s,k,H}^{j,s,h}(\tau,T) := \{ j \in \mathsf{R}_{f,s,H}(\tau,T) : \exp_f(\tau,T,j) = k \},$
- $\bullet \ \mathsf{R}^-_{f,s,k,H}(\tau,T) := \mathsf{R}^-(\tau,T) \cap \mathsf{R}_{f,s,k,H}(\tau,T), \\ \bullet \ \mathsf{R}^+_{f,s,k,H}(\tau,T) := \mathsf{R}^+(\tau,T) \cap \mathsf{R}_{f,s,k,H}(\tau,T).$

Maximal blocks of positions from $R(\tau, T)$ play an important role in our data structure. The starting positions of these blocks are defined as

$$R'(\tau, T) := \{ j \in R(\tau, T) : j - 1 \notin R(\tau, T) \}.$$

We also let:

- $R'^{-}(\tau, T) = R'(\tau, T) \cap R^{-}(\tau, T)$,
- $\bullet \ \mathsf{R}'^+(\tau,T) = \mathsf{R}'(\tau,T) \cap \mathsf{R}^+(\tau,T),$
- $\mathsf{R}_{f,H}^{\prime-}(\tau,T) = \mathsf{R}'(\tau,T) \cap \mathsf{R}_{f,H}^{-}(\tau,T),$ $\mathsf{R}_{f,H}^{\prime+}(\tau,T) = \mathsf{R}'(\tau,T) \cap \mathsf{R}_{f,H}^{\prime}(\tau,T).$

Lemma 8.48. Let $\tau \in [1...] \frac{n}{2}$ and f be any necklace-consistent function. For any position $j \in R(\tau,T)$ such that $j-1 \in R(\tau,T)$, it holds

- $\operatorname{root}_f(\tau, T, j-1) = \operatorname{root}_f(\tau, T, j),$
- $e(\tau, T, j-1) = e(\tau, T, j)$,
- $\begin{array}{l} \bullet \ \ \operatorname{tail}_f(\tau,T,j-1) = \operatorname{tail}_f(\tau,T,j), \\ \bullet \ \ e_f^{\mathrm{full}}(\tau,T,j-1) = e_f^{\mathrm{full}}(\tau,T,j), \\ \bullet \ \ \operatorname{type}(\tau,T,j-1) = \operatorname{type}(\tau,T,j). \end{array}$

Proof. Denote $P = T[j-1...j+3\tau-2)$, $P' = T[j...j+3\tau-1)$, p = per(P) and p' = per(P'). We first prove that p = p'. Suppose $p \neq p'$. Denote $X = T[j ... j + \tau)$. Observe that since P and P' overlap by $3\tau - 2 \geq \tau$ symbols, X occurs in both P and P'. Thus, X has periods p and p'. Note that it is not possible that $p \mid p'$, since this implies that the length-p' prefix of X is not primitive (contradicting p' = per(P')). By $p, p' \leq \frac{1}{3}\tau$ and the Weak Periodicity Lemma [FW65], we therefore have that X has period $p'' = \gcd(p, p')$. Since p does not divide p' we thus have p'' < p'. By definition, we also have $p'' \mid p$. This, however, again implies that Y[1...p'] is not primitive, contradicting $p' = \operatorname{per}(P')$. Thus, p = p'. We now show $\operatorname{root}_f(\tau, T, j - 1) = \operatorname{root}_f(\tau, T, j)$. Observe that $p, p' \leq \frac{1}{3}\tau$ implies that $\{T[j-1+\delta\mathinner{.\,.} j-1+\delta+p): \delta\in[0\mathinner{.\,.} p)\}=\{T[j+\delta\mathinner{.\,.} j+\delta+p): \delta\in[0\mathinner{.\,.} p)\}$. Thus, the strings T[j-1...j-1+p) and T[j...j+p) are cyclically equivalent. Consequently, since f is necklace-consistent, we have $\operatorname{root}_f(\tau, T, j-1) = f(T[j-1...j-1+p)) = f(T[j...j+p)) = \operatorname{root}_f(\tau, T, j)$.

By the above, it holds T[j-1] = T[j-1+p]. Thus, $e(\tau, T, j-1) = j-1+p + \text{LCE}_T(j-1, j-1+p) = j-1+p + \text{LCE}_T(j-1, j-1+p)$ $j + p + LCE_T(j, j + p) = e(\tau, T, j).$

Let H' (resp. H'') be a suffix (resp. prefix) of H of length head $f(\tau, T, j-1)$ (resp. tail $f(\tau, T, j-1)$). Then, $T[j-1...e(\tau,T,j-1)) = H'H'H''$, where $k = \exp_f(\tau,T,j-1)$. By $e(\tau,T,j-1) = e(\tau,T,j)$, and the uniqueness of this decomposition, it thus follows that we either have $T[j...e(\tau,T,j)) = H'[2...|H'|]H^kH''$ (when $|H'| \ge 1$), or $T[j ... e(\tau, T, j)) = H[2 ... p]H^{k-1}H''$ (if |H'| = 0). In either case, we have $tail_f(\tau, T, j) =$ $|H''| = \operatorname{tail}_f(\tau, T, j - 1).$

By the above, and the definition of $e_f^{\text{full}}(\tau, T, j-1)$ it follows that $e_f^{\text{full}}(\tau, T, j-1) = e(\tau, T, j-1) - \text{tail}_f(\tau, T, j-1) = e(\tau, T, j) - \text{tail}_f(\tau, T, j) = e_f^{\text{full}}(\tau, T, j)$.

By $e(\tau, T, j - 1) = e(\tau, T, j)$ and p = p', we immediately obtain $type(\tau, T, j - 1) = type(\tau, T, j)$.

Lemma 8.49. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$, and $j \in R(\tau, T)$. Then:

- 1. It holds $e(\tau, T, j) = \max\{j' \in [j ... n] : [j ... j'] \subseteq R(\tau, T)\} + 3\tau 1$, 2. If holds $b(\tau, T, j) = \min\{j' \in [1 ... j] : [j' ... j] \subseteq R(\tau, T)\}.$
- *Proof.* 1. Let $p = per(T[j ... j + 3\tau 1))$ and $j_e = e(\tau, T, j) 3\tau + 1$. Recall that, by definition, $p \leq \frac{1}{3}\tau$. On the other hand, by $e(\tau, T, j) = j + p + \text{LCE}_T(j, j + p)$, the string $T[j ... e(\tau, T, j)) = T[j ... j_e + 3\tau - 1)$ has period p. Thus, for every $j' \in [j ... j_e]$, it holds $\operatorname{per}(T[j' ... j' + 3\tau - 1)) \leq p \leq \frac{1}{3}\tau$. Hence $[j ... j_e] \subseteq \mathsf{R}(\tau, T)$. Moreover, by Lemma 8.48, for every $j' \in [j ... j_e]$, it holds $e(\tau, T, j') = e(\tau, T, j) = j_e$. Suppose now that $j_e+1 \in \mathsf{R}(\tau,T)$. By definition, we have $e(\tau,T,j_e+1) \geq (j_e+1)+3\tau-1=j_e+3\tau$. By Lemma 8.48, however, we would then also have $e(\tau, T, j_e + 1) = e(\tau, T, j_e) = j_e + 3\tau - 1$, a contradiction. Thus, $j_e + 1 \notin R(\tau, T)$, and hence $j_e = \max\{j' \in [j ... n] : [j ... j'] \subseteq \mathsf{R}(\tau, T)\}$. Consequently, $e(\tau, T, j) = j_e + 3\tau - 1 = \max\{j' \in [j ... n] : j' \in [j ... n] : j'$ $[j ... j'] \subseteq \mathsf{R}(\tau, T)\} + 3\tau - 1.$
- 2. Denote $j_b = b(\tau, T, j)$. By $b(\tau, T, j) = j \operatorname{lcs}(T[1...j], T[1...j + p))$, it follows that the string $T[j_b ... j + 3\tau - 1)$ has period p. Thus, for every $j' \in [j_b ... j]$, it holds $per(T[j'... j' + 3\tau - 1)) \leq p \leq p$ $\frac{1}{3}\tau$. Hence, $[j_b ... j] \subseteq \mathsf{R}(\tau, T)$. Suppose now that $j_b - 1 \in \mathsf{R}(\tau, T)$. By Lemma 8.48, we then have $per(T[j_b - 1...j_b + 3\tau - 2)) = |root_f(\tau, T, j_b - 1)| = |root_f(\tau, T, j_b)| = per(T[j_b ...j_b + 3\tau - 1)) = p. \text{ In}$ particular, $T[j_b-1]=T[j_b-1+p]$. This implies $j-\operatorname{lcs}(T[1...j),T[1...j+p))\leq j_b-1$, a contradiction. Thus, $b(\tau, T, j) = j_b = \min\{j' \in [1 ... j] : [j' ... j] \subseteq R(\tau, T)\}.$

Lemma 8.50. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$, f be any necklace-consistent function, and $j \in [1 ... n]$.

- 1. Let $P \in \Sigma^+$ be a τ -periodic pattern. Then, the following conditions are equivalent:
 - $j \in \operatorname{Occ}_{3\tau-1}(P,T)$,
 - $lcp(P, T[j ... n]) \ge 3\tau 1$,
 - $j \in \mathsf{R}(\tau, T)$, $\mathrm{root}_f(\tau, T, j) = \mathrm{root}_f(\tau, P)$, and $\mathrm{head}_f(\tau, T, j) = \mathrm{head}_f(\tau, P)$.

Moreover, if, letting $t = e(\tau, P) - 1$, it holds lep(P, T[j ... n]) > t, then:

- $e(\tau, P) 1 = e(\tau, T, j) j$,
- $$\begin{split} \bullet \ & \operatorname{tail}_f(\tau,P) = \operatorname{tail}_f(\tau,T,j), \\ \bullet \ & e_f^{\operatorname{full}}(\tau,P) 1 = e_f^{\operatorname{full}}(\tau,T,j) j, \end{split}$$
- $\exp_f(\tau, P) = \exp_f(\tau, T, j),$
- $type(\tau, P) = type(\tau, T, j)$.
- 2. Let $j' \in R(\tau, T)$. Then, the following conditions are equivalent:
 - $j \in \operatorname{Occ}_{3\tau-1}(j',T)$,
 - LCE_T $(j', j) \ge 3\tau 1$,
 - $j \in \mathsf{R}(\tau, T)$, $\mathrm{root}_f(\tau, T, j) = \mathrm{root}_f(\tau, T, j')$, and $\mathrm{head}_f(\tau, T, j') = \mathrm{head}_f(\tau, T, j)$.

Moreover, if letting $t = e(\tau, T, j) - j$, it holds LCE_T(j, j') > t, then:

```
 \begin{split} \bullet & \ e(\tau,T,j') - j' = e(\tau,T,j) - j, \\ \bullet & \ \tan\!f(\tau,T,j') = \tan\!f_f(\tau,T,j), \\ \bullet & \ e_f^{\rm full}(\tau,T,j') - j' = e_f^{\rm full}(\tau,T,j) - j, \\ \bullet & \ \exp_f(\tau,T,j') = \exp_f(\tau,T,j), \end{split}
```

• $\operatorname{type}(\tau, T, j') = \operatorname{type}(\tau, T, j)$.

Proof. 1. We begin by showing the three implications:

- Assume $j \in \text{Occ}_{3\tau-1}(P,T)$. Note that P being τ -periodic implies that $|P| \geq 3\tau 1$. Thus, $j \in \text{Occ}_{3\tau-1}(P,T)$ by definition implies $\text{lcp}(P,T[j ...n]) \geq \min(|P|,3\tau-1) \geq 3\tau-1$.
- Let us now assume $lcp(P, T[j ...n]) \ge 3\tau 1$. This implies $j \in [1..n 3\tau + 2]$. Moreover, denoting P' = T[j ...n], it then follows by Lemma 8.45, that P' is τ -periodic and it holds $root_f(\tau, P') = root_f(\tau, P)$ and $head_f(\tau, P') = head_f(\tau, P)$. Consequently, $per(T[j ...j + 3\tau 1)) = per(P'[1...3\tau 1]) \le \frac{1}{3}\tau$, i.e., $j \in R(\tau, T)$. We then also immediately obtain $root_f(\tau, T, j) = root_f(\tau, P') = root_f(\tau, P)$ and $head_f(\tau, T, j) = head_f(\tau, P') = head_f(\tau, P)$.
- Let us finally assume that it holds $j \in R(\tau, T)$, $\operatorname{root}_f(\tau, T, j) = \operatorname{root}_f(\tau, P)$, and $\operatorname{head}_f(\tau, T, j) = \operatorname{head}_f(\tau, P)$. Denote P' = T[j ...n]. Note that $j \in R(\tau, T)$ implies that P' is τ -periodic. Moreover, by definition, we have $\operatorname{root}_f(\tau, P') = \operatorname{root}_f(\tau, T, j)$ and $\operatorname{head}_f(\tau, P') = \operatorname{head}_f(\tau, T, j)$. Thus, we obtain $\operatorname{root}_f(\tau, P') = \operatorname{root}_f(\tau, P)$ and $\operatorname{head}_f(\tau, P') = \operatorname{head}_f(\tau, P)$. Consequently, it follows by Lemma 8.45, that $\operatorname{lcp}(P, P') \geq 3\tau 1$. By definition of P', we thus obtain $j \in \operatorname{Occ}_{3\tau 1}(P, T)$.

Let us now assume lcp(P, T[j ... n]) > t, where $t = e(\tau, P) - 1$. Note that this implies $|P| \ge lcp(P, P') \ge t + 1 = e(\tau, P)$. Letting P' = T[j ... n], we also have $lcp(P, P') \ge t + 1 = e(\tau, P)$. Thus, by combining Lemma 8.45 with the definitions of the values below, we obtain:

```
• e(\tau, P) - 1 = e(\tau, P') - 1 = e(\tau, T, j) - j,

• tail_f(\tau, P) = tail_f(\tau, P') = tail_f(\tau, T, j),
```

- $e_f^{\text{full}}(\tau, P) 1 = e_f^{\text{full}}(\tau, P') 1 = e_f^{\text{full}}(\tau, T, j) j$
- $\exp_f(\tau, P) = \exp_f(\tau, P') = \exp_f(\tau, T, j),$
- $\operatorname{type}(\tau, P) = \operatorname{type}(\tau, P') = \operatorname{type}(\tau, T, j)$.
- 2. Denote P:=T[j'..n]. Observe that $j' \in \mathbb{R}(\tau,T)$ implies that P is τ -periodic and, by definition, it holds $\operatorname{Occ}_{3\tau-1}(j',T) = \operatorname{Occ}_{3\tau-1}(P,T)$, $\operatorname{LCE}_T(j',j) = \operatorname{lcp}(P,T[j..n])$, $\operatorname{root}_f(\tau,T,j') = \operatorname{root}_f(\tau,P)$, $\operatorname{head}_f(\tau,T,j') = \operatorname{head}_f(\tau,P)$, $e(\tau,T,j') j' = e(\tau,P) 1$, $\operatorname{tail}_f(\tau,T,j') = \operatorname{tail}_f(\tau,P)$, $e_f^{\operatorname{full}}(\tau,T,j') j' = e_f^{\operatorname{full}}(\tau,P) 1$, $\operatorname{exp}_f(\tau,T,j') = \exp_f(\tau,P)$, and $\operatorname{type}(\tau,T,j') = \operatorname{type}(\tau,P)$. Thus, all claims follow by Lemma 8.50(1).

Lemma 8.51. Let $\tau \in [1..\lfloor \frac{n}{2} \rfloor]$, f be any necklace-consistent function, and $j \in R(\tau, T)$. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{root}_f(\tau, P) = \operatorname{root}_f(\tau, T, j)$ and $\operatorname{head}_f(\tau, P) = \operatorname{head}_f(\tau, T, j)$. Then, letting $t_1 = e(\tau, T, j) - j$ and $t_2 = e(\tau, P) - 1$, it holds $\operatorname{lcp}(T[j ... n], P) \ge \min(t_1, t_2)$ and:

```
1. If \operatorname{type}(\tau, T, j) \neq \operatorname{type}(\tau, P) or t_1 \neq t_2, then T[j..n] \neq P and \operatorname{lcp}(T[j..n], P) = \min(t_1, t_2),
```

- $2. \ \ \textit{If} \ \mathsf{type}(\tau, T, j) \neq \mathsf{type}(\tau, P), \ \textit{then} \ T[j\mathinner{\ldotp\ldotp} n] \prec P \ \textit{if and only if} \ \mathsf{type}(\tau, T, j) < \mathsf{type}(\tau, P),$
- 3. If $type(\tau, T, j) = -1$, then $t_1 < t_2$ implies $T[j ... n] \prec P$,
- 4. If type $(\tau, T, j) = +1$, then $t_1 < t_2$ implies T[j ... n] > P,
- 5. If $\operatorname{type}(\tau, T, j) = \operatorname{type}(\tau, P) = -1$, and $t_1 \neq t_2$, then $t_1 < t_2$ if and only if $T[j ... n] \prec P$,
- 6. If $\operatorname{type}(\tau, T, j) = \operatorname{type}(\tau, P) = +1$, and $t_1 \neq t_2$, then $t_1 < t_2$ if and only if T[j ... n] > P.

Proof. Denote $P_1 := T[j ... n]$ and $P_2 := P$. Observe that P_1 is τ -periodic and it holds $\operatorname{root}_f(\tau, P_1) = \operatorname{root}_f(\tau, T, j) = \operatorname{root}_f(\tau, P_2)$ and $\operatorname{head}_f(\tau, P_1) = \operatorname{head}_f(\tau, T, j) = \operatorname{head}_f(\tau, P_2)$. Moreover, we then have $\operatorname{type}(\tau, P_1) = \operatorname{type}(\tau, T, j)$ and $e(\tau, P_2) - 1 = e(\tau, T, j) - j = t_1$. All claims thus follow by Lemma 8.46. \square

Lemma 8.52. Let $\tau \in [1..\lfloor \frac{n}{2} \rfloor]$, f be any necklace-consistent function, and $j \in R(\tau, T)$. Let $j' \in R(\tau, T)$ be such that $\text{root}_f(\tau, T, j') = \text{root}_f(\tau, T, j)$ and $\text{head}_f(\tau, T, j') = \text{head}_f(\tau, T, j)$. Then, letting $t_1 = e(\tau, T, j) - j$ and $t_2 = e(\tau, T, j') - j'$, it holds $\text{LCE}_T(j, j') \geq \min(t_1, t_2)$ and:

1. If $\operatorname{type}(\tau, T, j) \neq \operatorname{type}(\tau, T, j')$ or $t_1 \neq t_2$, then $\operatorname{LCE}_T(j, j') = \min(t_1, t_2)$,

- 2. If $\operatorname{type}(\tau, T, j) \neq \operatorname{type}(\tau, T, j')$, then $T[j \dots n] \prec T[j' \dots n]$ if and only if $\operatorname{type}(\tau, T, j) < \operatorname{type}(\tau, T, j')$,
- 3. If $\operatorname{type}(\tau, T, j) = \operatorname{type}(\tau, T, j') = -1$ and $t_1 \neq t_2$, then $t_1 < t_2$ if and only if $T[j \dots n] \prec T[j' \dots n]$,
- 4. If $\operatorname{type}(\tau, T, j) = \operatorname{type}(\tau, T, j') = +1$ and $t_1 \neq t_2$, then $t_1 < t_2$ if and only if $T[j ... n] \succ T[j' ... n]$.

Proof. Denote $P_3 := T[j' ... n]$. Observe that P_3 is τ -periodic and it holds $\operatorname{root}_f(\tau, P_3) = \operatorname{root}_f(\tau, T, j') =$ $\operatorname{root}_f(\tau, T, j)$ and $\operatorname{head}_f(\tau, P_3) = \operatorname{head}_f(\tau, T, j') = \operatorname{head}_f(\tau, T, j)$. Moreover, we then have $\operatorname{type}(\tau, P_3) = \operatorname{head}_f(\tau, T, j')$ type (τ, T, j') and $e(\tau, P_3) - 1 = e(\tau, T, j') - j' = t_2$. All claims thus follow by Lemma 8.51.

Lemma 8.53. Let $\tau \in [1..\lfloor \frac{n}{2} \rfloor]$. Let $j, j', j'' \in [1..n]$ be such that $j, j'' \in \mathsf{R}(\tau, T), \ j' \not\in \mathsf{R}(\tau, T), \ and <math>j < j' < j''$. Then, it holds $e(\tau, T, j) \leq j'' + \tau - 1$.

Proof. Consider any $i \in R(\tau, T)$ and let $b_i = b(\tau, T, i)$, $e_i = e(\tau, T, i)$, and $p_i = per(T[i ... i + 3\tau - 1))$. By definition, the fragment $T[b_e \dots e_i]$ has a period p. Since, however, it contains $T[i \dots i + 3\tau - 1]$ as a substring, we must have $per(T[b_i ... e_i)) = p_i$. On the other hand, by definition of $b(\tau, T, i)$ and $e(\tau, T, i)$, the substring $T[b_i \dots e_i]$ cannot be extended in either left or right in T without increasing its shortest period. Thus, by $p \leq \lfloor \frac{1}{3}\tau \rfloor$, $T[b_i \ldots e_i)$ is a therefore a run [KK99, Mai89]. Any two distinct runs with periods at most k must overlap by less than 2k symbols (see, e.g., [Koc18, Fact 2.2.4]). Lastly, observe that by Lemma 8.49, we then also have $[b_i ... e_i - 3\tau + 1] \subseteq \mathsf{R}(\tau, T), \ b_i - 1 \not\in \mathsf{R}(\tau, T), \ \mathrm{and} \ e_i - 3\tau + 2 \not\in \mathsf{R}(\tau, T).$

Let us now consider substrings $T[b(\tau,T,j) ... e(\tau,T,j))$ and $T[b(\tau,T,j'') ... e(\tau,T,j''))$. By the above discussion we have $e(\tau, T, j) - b(\tau, T, j'') < 2\lfloor \frac{1}{3}\tau \rfloor$. By $b(\tau, T, j'') \leq j''$, we thus obtain $e(\tau, T, j) \leq b(\tau, T, j'') + b(\tau, T, j'') \leq b(\tau, T,$ $2\lfloor \frac{1}{3}\tau \rfloor \le j'' + \tau - 1.$

Lemma 8.54. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. For any $j, j' \in \mathsf{R}'(\tau, T), \ j \neq j'$ implies $e_f^{\mathrm{full}}(\tau, T, j) \neq e_f^{\mathrm{full}}(\tau, T, j')$.

Proof. Without the loss of generality, let us assume j < j'. We first derive an upper bound on $e_f^{\text{full}}(\tau, T, j)$ and a lower bound on $e_f^{\text{full}}(\tau, T, j')$:

- By definition, it holds $j'-1 \notin \mathsf{R}(\tau,T)$. By Lemma 8.53 applied for j, j'-1, and j', we thus obtain
- by definition, it holds f = 1 ∉ R(f, T). By Lemma 8.33 applied for f, f = 1, and f, we thus obtain e(τ, T, j) ≤ j' + τ 1. Consequently, e^{full}_f(τ, T, j) ≤ e(τ, T, j) ≤ j' + τ 1.
 By definition, e(τ, T, j') j' ≥ 3τ 1, or equivalently, e(τ, T, j') ≥ j' + 3τ 1. On the other hand, e(τ, T, j') e^{full}_f(τ, T, j') = tail_f(τ, T, j') ≤ ⌊¹₃τ⌋. Combining the two inequalities, we thus obtain e^{full}_f(τ, T, j') ≥ e(τ, T, j') ⌊¹₃τ⌋ ≥ j' + 3τ 1 ⌊¹₃τ⌋ ≥ j' + 2τ 1.

By combining the two bounds, we obtain $e_f^{\text{full}}(\tau, T, j) \leq j' + \tau - 1 < j' + 2\tau - 1 \leq e_f^{\text{full}}(\tau, T, j')$. Thus, $e_f^{\text{full}}(\tau, T, j) \neq e_f^{\text{full}}(\tau, T, j')$.

Lemma 8.55. Let $\tau \in [1... | \frac{n}{2} |]$ and f be any necklace-consistent function. Let $i \in \mathbb{R}(\tau, T)$ and $\ell = e(\tau, T, i) - i$. For every $j' \in [1..n]$, $\delta \in [0..\ell]$, and $\ell' > \ell$, $T^{\infty}[j' - \delta..j' - \delta + \ell') = T^{\infty}[i..i + \ell')$ implies that, letting $j = j' - \delta$, it holds $j \in R(\tau, T)$ and:

- $\operatorname{root}_f(\tau, T, j) = \operatorname{root}_f(\tau, T, i)$,
- head $f(\tau, T, j) = \text{head } f(\tau, T, i)$,
- $e(\tau, T, j) j = e(\tau, T, i) i$,
- $tail_f(\tau, T, j) = tail_f(\tau, T, i),$
- $\begin{array}{l} \bullet \ e_f^{\mathrm{full}}(\tau,T,j) j = e_f^{\mathrm{full}}(\tau,T,i) i, \\ \bullet \ \exp_f(\tau,T,j) = \exp_f(\tau,T,i), \end{array}$
- $type(\tau, T, j) = type(\tau, T, i)$.

Proof. First, observe that if $p \in [1..n]$ and for $t \geq 0$ (resp. $t' \geq 0$) the substring $T^{\infty}[p-t..p]$ (resp. $T^{\infty}[p ... p + t')$ does not contain symbol T[n], then it holds $p - t \ge 1$ (resp. $p + t' \le n$). Since for every $j \in \mathsf{R}(\tau,T)$, we have $e(\tau,T,j) \leq n$, then $i+\ell \leq n$ and by the uniqueness of T[n] in T, the substring $T^{\infty}[i\ldots i+\ell]$ does not contain T[n]. Consequently, by $0 \le \delta \le \ell < \ell'$ and $T^{\infty}[j' - \delta ... j' - \delta + \ell') = T^{\infty}[i ... i + \ell')$, we have $T^{\infty}[i-\delta ...i) = T^{\infty}[j'-\delta ...j')$ and $T^{\infty}[i...i+\delta'] = T^{\infty}[j'...j'+\delta']$ (where $\delta' = \ell - \delta$). Thus, $j' - \delta \ge 1$ and $j + \ell = j' + \delta' \le n$, i.e., $[j ... j + \ell] \subseteq [1 ... n]$. Thus, $LCE_T(i, j) \ge \ell + 1 > e(\tau, T, i) - i \ge 3\tau - 1$. The lemma thus follows by Lemma 8.50(2).

Lemma 8.56. Let $\tau \in [1...] \frac{n}{2}$ and f be any necklace-consistent function. Let $j \in \mathbb{R}(\tau,T)$. Consider any $t > 3\tau - 1$.

- 1. Let $P \in \Sigma^+$ be a τ -periodic pattern satisfying $e(\tau, P) \leq |P|$. If $j \in Occ_t(P, T)$, then it holds:

 - $\begin{array}{l} \bullet \ \operatorname{exp}_f^{\operatorname{cut}}(\tau,T,j,t) = \operatorname{exp}_f^{\operatorname{cut}}(\tau,P,t), \\ \bullet \ e_f^{\operatorname{cut}}(\tau,T,j,t) j = e_f^{\operatorname{cut}}(\tau,P,t) 1. \end{array}$
- 2. Let $j' \in R(\tau, T)$. If $j \in Occ_t(j', T)$, then it holds:

 - $\begin{array}{l} \bullet \ \operatorname{exp}_f^{\operatorname{cut}}(\tau,T,j,t) = \operatorname{exp}_f^{\operatorname{cut}}(\tau,T,j',t), \\ \bullet \ e_f^{\operatorname{cut}}(\tau,T,j,t) j = e_f^{\operatorname{cut}}(\tau,T,j',t) j'. \end{array}$

Proof. Denote $H = \text{root}_f(\tau, T, j)$, $s = \text{head}_f(\tau, T, j)$, and p = |H|.

- 1. We first observe that $t \geq 3\tau 1$ implies that $j \in \text{Occ}_t(P,T) \subseteq \text{Occ}_{3\tau 1}(P,T)$. Thus, by Lemma 8.50(1), it follows that root $f(\tau, P) = H$ and head $f(\tau, P) = s$. Denote $t' = e(\tau, P) - 1$. We consider two cases:
 - Assume $t \leq t'$. Then, it holds $\exp_f^{\text{cut}}(\tau, P, t) = \min(\exp_f(\tau, P), \lfloor \frac{t-s}{p} \rfloor) = \min(\lfloor \frac{e(\tau, P) 1 s}{p} \rfloor, \lfloor \frac{t-s}{p} \rfloor) = \min(\lfloor \frac{t'-s}{p} \rfloor, \lfloor \frac{t-s}{p} \rfloor) = \lfloor \frac{t-s}{p} \rfloor$. Note that we then also have $|P| \geq e(\tau, P) 1 = t' \geq t$. By definition of $\operatorname{Occ}_t(P, T, t)$, this implies $\operatorname{lcp}(P, T[j \dots n]) \geq \min(|P|, t) = t$. By definition of $e(\tau, P)$, we therefore have $p + \operatorname{lcp}(P[1 ... |P|], P[1 + p ... |P|]) = e(\tau, P) - 1 \ge t$. Consequently, the assumption $\operatorname{lcp}(P, T[j ... n]) \ge t$ implies $p + \text{LCE}_T(j, j + p) \ge t$, and hence $e(\tau, T, j) = j + p + \text{LCE}_T(j, j + p) \ge j + t$, or equivalently, $e(\tau, T, j) - j \ge t$. This implies $\exp_f(\tau, T, j) = \lfloor \frac{e(\tau, T, j) - j - s}{p} \rfloor \ge \lfloor \frac{t - s}{p} \rfloor$. We thus obtain $\exp_f^{\text{cut}}(\tau, T, j, t) = \min(\exp_f(\tau, T, j), \lfloor \frac{t - s}{p} \rfloor) = \lfloor \frac{t - s}{p} \rfloor = \exp_f^{\text{cut}}(\tau, P, t)$.

 • Let us now assume t > t'. Recall that we assumed $|P| \ge e(\tau, P) = t' + 1$. Consequently, from
 - $j \in \operatorname{Occ}_t(P,T)$, it follows that $\operatorname{lcp}(P,T[j\mathinner{\ldotp\ldotp\ldotp} n]) \ge \min(|P|,t) > t'$. By Lemma 8.50(1), it follows that $e(\tau,T,j)-j=e(\tau,P)-1=t'$. This in turn implies that $\exp_f(\tau,T,j)=\lfloor\frac{e(\tau,T,j)-j-s}{p}\rfloor=\lfloor\frac{e(\tau,P)-1-s}{p}\rfloor=\exp_f(\tau,P)$. Hence, we obtain $\exp_f^{\operatorname{cut}}(\tau,T,j,t)=\min(\exp_f(\tau,T,j),\lfloor\frac{t-s}{p}\rfloor)=\min(\exp_f(\tau,P),\lfloor\frac{t-s}{p}\rfloor)=\exp_f^{\operatorname{cut}}(\tau,P,t)$.

In both cases, we thus have $\exp_f^{\mathrm{cut}}(\tau,T,j,t) = \exp_f^{\mathrm{cut}}(\tau,P,t)$, i.e., the first claim. This immediately gives the second claim, since, $e_f^{\mathrm{cut}}(\tau,T,j,t) - j = s + \exp_f^{\mathrm{cut}}(\tau,T,j,t) \cdot p = s + \exp_f^{\mathrm{cut}}(\tau,P,t) \cdot p = e_f^{\mathrm{cut}}(\tau,P,t) - 1$.

2. Let P = T[j' ... n]. Note that P is τ -periodic and, by the uniqueness of T[n] in T, it holds $|P| \ge e(\tau, P)$ and $j \in \mathrm{Occ}_t(j',T) = \mathrm{Occ}_t(P,T)$. Note also that, by definition, it holds $\exp_f^{\mathrm{cut}}(\tau,T,j',t) = \exp_f^{\mathrm{cut}}(\tau,P,t)$ and $e_f^{\text{cut}}(\tau, T, j', t) - j' = e_f^{\text{cut}}(\tau, P, t) - 1$. Thus, the claims follow by Lemma 8.56(1).

Lemma 8.57. Let $\tau \in [1, \lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Let $j \in R(\tau, T)$. Denote $s = \text{head}_f(\tau, T, j), H = \text{root}_f(\tau, T, j), p = |H|, H' = H(p - s ... p].$ Let t, t' be such that $3\tau - 1 \le t' \le t$. Let P be a length-t prefix of $H'H^{\infty}$. Denote $k = \lfloor \frac{t-s}{p} \rfloor$, $k' = \lfloor \frac{t'-s}{p} \rfloor$, $x_l = s + k'p$, and $y_u = T[j+s \ldots j+s+t'-x_l)$. Then:

- P is τ -periodic,
- P does not contain T[n],
- $j \in \operatorname{Occ}_{3\tau-1}(P,T)$,
- head $f(\tau, P) = s$,
- $\operatorname{root}_f(\tau, P) = H$,
- $e(\tau, P) 1 = t$,
- $type(\tau, P) = -1$,
- $\exp_f(\tau, P) = k$,

- $\exp_f^{\text{cut}}(\tau, P, t') = k',$ $e_f^{\text{cut}}(\tau, P, t') 1 = x_l,$ $P[e_f^{\text{cut}}(\tau, P, t') \dots t'] = y_u.$

Proof. By $|P| = t \ge t' \ge 3\tau - 1$ and $|H| \le \frac{1}{3}\tau$, it holds $\operatorname{per}(P[1..3\tau - 1]) \le |H| \le \frac{1}{3}\tau$. Thus, P is τ -periodic (Definition 8.1). Since P is a substring of H^{∞} , P does not contain symbol T[n]. Next, observe that $\operatorname{root}_f(\tau,T,j)=H$ and $\operatorname{head}_f(\tau,T,j)=s$ implies that, letting P' be a prefix of $H'H^{\infty}$ of length $3\tau-1,T[j..n]$

has P' as a prefix. Thus, $lcp(P, T[j ... n]) \ge lcp(P, P') = 3\tau - 1$, i.e., $j \in Occ_{3\tau - 1}(P, T)$. By Lemma 8.50(1), this also implies $head_f(\tau, P) = s$ and $root_f(\tau, P) = H$. This in turn implies $e(\tau, P) - 1 = p + lcp(P, P[1 + p . t]) = t$. Consequently, $type(\tau, P) = -1$. Next, observe that $exp_f(\tau, P) = \lfloor \frac{e(\tau, P) - 1 - s}{p} \rfloor = \lfloor \frac{t - s}{p} \rfloor = k$. We then also have $exp_f^{cut}(\tau, P, t') = min(exp_f(\tau, P), \lfloor \frac{t' - s}{p} \rfloor) = min(k, k') = k'$, and $e_f^{cut}(\tau, P, t') - 1 = s + k'p = x_l$. Lastly, observe that $P[e_f^{cut}(\tau, P, t') . . t']$ is a prefix of H of length $t' - e_f^{cut}(\tau, P, t') + 1 = t' - x_l$. By T[j + s . . j + s + p) = H we thus have $P[e_f^{cut}(\tau, P, t') . . t'] = T[j + s . . j + s + t' - x_l) = y_u$.

Lemma 8.58. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Let $H \in \Sigma^+$, p = |H|, $s \in [0 ... p)$, and $k_{\min} = \lceil \frac{3\tau - 1 - s}{p} \rceil - 1$. Then:

1. For every $k \in [1..k_{\min}]$,

$$\mathsf{R}^-_{f,s,k,H}(\tau,T) \subseteq \{e_f^{\mathrm{full}}(\tau,T,j) - s - kp : j \in \mathsf{R}'^-_{f,H}(\tau,T) \ \ and \ s + kp \leq e_f^{\mathrm{full}}(\tau,T,j) - j\},$$

2. For every $k \in (k_{\min} ... n]$,

$$\mathsf{R}^-_{f,s,k,H}(\tau,T) = \{e_f^{\mathrm{full}}(\tau,T,j) - s - kp : j \in \mathsf{R}'^-_{f,H}(\tau,T) \ \ and \ s + kp \leq e_f^{\mathrm{full}}(\tau,T,j) - j\}.$$

Proof. For every $k \in [1 ... n]$, denote $A_k := \{e_f^{\text{full}}(\tau, T, j) - s - kp : j \in \mathsf{R}'_{f,H}(\tau, T) \text{ and } s + kp \leq e_f^{\text{full}}(\tau, T, j) - j\}$. Consider any $k \in [1 ... n]$. In the first step, we prove that $\mathsf{R}^-_{f,s,k,H}(\tau, T) \subseteq A_k$. Let $j \in \mathsf{R}^-_{f,s,k,H}(\tau, T)$, and $j' = b(\tau, T, j)$. Then, by Lemma 8.49(2), it holds $[j' ... j] \subseteq \mathsf{R}(\tau, T)$. By Lemma 8.48 we thus have $\operatorname{root}_f(\tau,T,j') = \operatorname{root}_f(\tau,T,j) = H$ and $\operatorname{type}(\tau,T,j') = \operatorname{type}(\tau,T,j) = -1$. On the other hand, by Lemma 8.49(2), $j' \in R'(\tau,T)$. Hence, $j' \in R'_{f,H}(\tau,T)$. Note that Lemma 8.48 also implies $e_f^{\operatorname{full}}(\tau,T,j') = e_f^{\operatorname{full}}(\tau,T,j)$. Consequently, since $j' \leq j$, it holds $e_f^{\operatorname{full}}(\tau,T,j') - j' \geq e_f^{\operatorname{full}}(\tau,T,j') - j = e_f^{\operatorname{full}}(\tau,T,j) + \exp_f(\tau,T,j)|\operatorname{root}_f(\tau,T,j)| = s + kp$. On the other hand, $e_f^{\operatorname{full}}(\tau,T,j) - j = s + kp$ and $e_f^{\operatorname{full}}(\tau,T,j) = e_f^{\operatorname{full}}(\tau,T,j')$ imply $j = e_f^{\operatorname{full}}(\tau,T,j) - s - kp = e_f^{\operatorname{full}}(\tau,T,j') - s - kp$. We have thus proved that $e_f^{\operatorname{full}}(\tau,T,j') = e_f^{\operatorname{full}}(\tau,T,j')$ imply $f_f^{\operatorname{full}}(\tau,T,j') = e_f^{\operatorname{full}}(\tau,T,j') - s - kp$.

Let us now consider $k \in (k_{\min} ... n]$. In the second step, we prove that for such k, it holds $A_k \subseteq \mathsf{R}^-_{f,s,k,H}(\tau,T)$. Let $j \in A_k$. Then, there exists $j' \in \mathsf{R}'_{f,H}(\tau,T)$ such that $s+kp \leq e_f^{\mathrm{full}}(\tau,T,j')-j'$ and $j=e_f^{\mathrm{full}}(\tau,T,j')-s-kp$. The assumptions immediately imply that $j' \leq j$. On the other hand, $k > k_{\min}$ implies that $e(\tau,T,j')-j \geq e_f^{\mathrm{full}}(\tau,T,j')-j=s+kp \geq s+(k_{\min}+1)\cdot p=s+\lceil\frac{3\tau-1-s}{p}\rceil p \geq s+\frac{3\tau-1-s}{p}\cdot p=3\tau-1$. Thus, it follows by Lemma 8.49(1), that $[j'...j] \subseteq \mathsf{R}(\tau,T)$. By Lemma 8.48, we thus have $j \in \mathsf{R}^-_{f,H}(\tau,T)$ and $e_f^{\mathrm{full}}(\tau,T,j)=e_f^{\mathrm{full}}(\tau,T,j')$. Thus, head $f(\tau,T,j)=(e_f^{\mathrm{full}}(\tau,T,j)-j) \mod p=(e_f^{\mathrm{full}}(\tau,T,j')-j) \mod p=(s+kp) \mod p=s$ and $\exp_f(\tau,T,j)=\lfloor\frac{e_f^{\mathrm{full}}(\tau,T,j)-j}{p}\rfloor=\lfloor\frac{s+kp}{p}\rfloor=k$. We have thus proved that $j \in \mathsf{R}^-_{f,s,k,H}(\tau,T)$.

Lemma 8.59. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Let $H \in \Sigma^+$, p = |H|, $s \in [0 ... p)$, and $k_{\min} = \lceil \frac{3\tau - 1 - s}{p} \rceil - 1$. Then:

- $\begin{array}{l} \bullet \ \ For \ every \ k \in [1 \mathinner{.\,.} k_{\min}], \ |\mathsf{R}^-_{f,s,k,H}(\tau,T)| \leq |\{j \in \mathsf{R}'^-_{f,H}(\tau,T) : s + kp \leq e^{\mathrm{full}}_f(\tau,T,j) j\}|, \\ \bullet \ \ For \ every \ k \in (k_{\min} \mathinner{.\,.} n], \ |\mathsf{R}^-_{f,s,k,H}(\tau,T)| = |\{j \in \mathsf{R}'^-_{f,H}(\tau,T) : s + kp \leq e^{\mathrm{full}}_f(\tau,T,j) j\}|. \\ \end{array}$

Proof. First, observe that adding the same value to every element of a set does not change its cardinality. Moreover, by Lemma 8.54, for every $i, i' \in \mathsf{R}'^-(\tau, T), i \neq i'$ implies $e_f^{\mathrm{full}}(\tau, T, i) \neq e_f^{\mathrm{full}}(\tau, T, i')$. It thus follows by Lemma 8.58 that for $k \in [1..k_{\min}]$,

$$\begin{split} |\mathsf{R}^{-}_{f,s,k,H}(\tau,T)| &\leq |\{e^{\mathrm{full}}_{f}(\tau,T,j) - s - kp : j \in \mathsf{R}'^{-}_{f,H}(\tau,T) \text{ and } s + kp \leq e^{\mathrm{full}}_{f}(\tau,T,j) - j\}| \\ &= |\{e^{\mathrm{full}}_{f}(\tau,T,j) : j \in \mathsf{R}'^{-}_{f,H}(\tau,T) \text{ and } s + kp \leq e^{\mathrm{full}}_{f}(\tau,T,h) - j\}| \\ &= |\{j \in \mathsf{R}'^{-}_{f,H}(\tau,T) : s + kp \leq e^{\mathrm{full}}_{f}(\tau,T,h) - j\}|. \end{split}$$

The proof for $k \in (k_{\min} ... n]$ follows analogously, except the first inequality is replaced with an equality.

Definition 8.60. Let $\tau \in [1, \lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Let $H \in \Sigma^+$. We define

$$\mathrm{Seed}_{f,H}^-(\tau,T) := \{(e_f^{\mathrm{full}}(\tau,T,p), \min(7\tau,e_f^{\mathrm{full}}(\tau,T,p) - b(\tau,T,p))) : p \in \mathrm{comp}_{14\tau}(\mathsf{R}_{f,H}^-(\tau,T),T)\}.$$

Lemma 8.61. Let $\tau \in [1... \lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. For every $H \in \Sigma^+$, it holds

$$Seed_{f,H}^{-}(\tau,T) = \{(e_f^{\text{full}}(\tau,T,p_j), \min(7\tau, e_f^{\text{full}}(\tau,T,p_j) - b(\tau,T,p_j))) : j \in [1 ... m] \text{ and } p_j \in \mathsf{R}_{f,H}^{-}(\tau,T)\},$$

where Seed_{f,H}⁻ (τ,T) is as in Definition 8.60 and $(p_j,t_j)_{j\in[1..m]} = \mathcal{I}(\text{comp}_{14\tau}(\mathsf{R}(\tau,T),T)).$

Proof. Denote the set on the right-hand side by A. For every $p \in \mathsf{R}(\tau,T)$, let us also denote $\alpha(p) := e_f^{\mathrm{full}}(\tau,T,p)$ and $\beta(p) := \min(7\tau, e_f^{\mathrm{full}}(\tau,T,p) - b(\tau,T,p))$.

Let $a \in A$. Then, there exists $j \in [1..m]$ such that $p_j \in \mathsf{R}^-_{f,H}(\tau,T)$ and $a = (\alpha(p_j),\beta(p_j))$. Recall that $\operatorname{comp}_{14\tau}(\mathsf{R}(\tau,T),T) = \mathsf{R}(\tau,T) \cap \mathsf{C}(14\tau,T)$. By Definition 4.2, it holds $[p_j \dots p_j + t_j) \subseteq \mathsf{R}(\tau,T) \cap \mathsf{C}(14\tau,T)$. In particular, $p_j \in \mathsf{C}(14\tau,T)$. Combining with $p_j \in \mathsf{R}^-_{f,H}(\tau,T)$, we thus obtain $p_j \in \mathsf{R}^-_{f,H}(\tau,T) \cap \mathsf{C}(14\tau,T) = \operatorname{comp}_{14\tau}(\mathsf{R}^-_{f,H}(\tau,T),T)$. By Definition 8.60, we thus obtain $(\alpha(p_j),\beta(p_j)) \in \operatorname{Seed}^-_{f,H}(\tau,T)$. We have thus proved $a \in \operatorname{Seed}^-_{f,H}(\tau,T)$.

We now show the second inclusion. Let $a \in \operatorname{Seed}_{f,H}^-(\tau,T)$. Then, there exists $p \in \operatorname{comp}_{14\tau}(\mathsf{R}_{f,H}^-(\tau,T),T) = \mathsf{R}_{f,H}^-(\tau,T) \cap \mathsf{C}(14\tau,T)$ such that $a = (\alpha(p),\beta(p))$. Let $p' = \min\{t \in [1 \dots p] : [t \dots p] \subseteq \mathsf{R}(\tau,T) \cap \mathsf{C}(14\tau,T)\}$. We prove three properties of p':

- First, we observe that by Lemma 8.48, it follows that $\operatorname{type}(\tau, T, p') = \operatorname{type}(\tau, T, p)$ and $\operatorname{root}_f(\tau, T, p') = \operatorname{root}_f(\tau, T, p)$. Thus, $p' \in \mathsf{R}_{f,H}^-(\tau, T)$.
- Next, observe that, by definition, we have $p'-1 \notin \mathsf{R}(\tau,T) \cap \mathsf{C}(14\tau,T)$, i.e., p' is the leftmost element of a maximal block in $\mathsf{R}(\tau,T) \cap \mathsf{C}(14\tau,T)$. Thus, there exists $j \in [1...m]$ such that $p_j = p'$.
- Lastly, observe that by Lemmas 8.48 and 8.49, it holds $e_f^{\text{full}}(\tau, T, p') = e_f^{\text{full}}(\tau, T, p)$ and $b(\tau, T, p') = b(\tau, T, p)$. Consequently, $(\alpha(p'), \beta(p')) = (\alpha(p), \beta(p))$.

Putting everything together, we thus obtain that there exists $j \in [1..m]$ such that $p' = p_j$ and $p_j \in \mathsf{R}_{f,H}^-(\tau,T)$. Thus, $(\alpha(p'), \beta(p')) \in A$. By $(\alpha(p'), \beta(p')) = (\alpha(p), \beta(p)) = a$, we thus have $a \in A$.

Corollary 8.62. Let $\tau \in [1... \lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Then, it holds

$$\sum_{H \in \Sigma^+} |\mathrm{Seed}_{f,H}^-(\tau,T)| \leq |\mathcal{I}(\mathrm{comp}_{14\tau}(\mathsf{R}(\tau,T),T))|.$$

Proof. The result follows immediately from Lemma 8.61.

Lemma 8.63. Let $\tau \in [1..\lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. For every $H \in \Sigma^+$, the labels in WeightedIntervals_{7 τ}(Seed⁻_{f,H}(τ , T), T) (Definition 7.1) are unique.

Proof. For every $j \in \mathsf{R}(\tau,T)$, denote $\alpha(j) = e_f^{\mathrm{full}}(\tau,T,j)$ and $\beta(j) = \min(7\tau,e_f^{\mathrm{full}}(\tau,T,j) - b(\tau,T,j))$. Denote $\mathcal{I} = \mathrm{WeightedIntervals}_{7\tau}(\mathrm{Seed}_{f,H}^-(\tau,T),T)$. Let $q_1,q_2 \in \mathcal{I}$. For $k \in \{1,2\}$, denote $q_k = (e_k,w_k,\ell_k)$, and assume $\ell_1 = \ell_2$. We need to prove that $q_1 = q_2$. By Definition 7.1 and the definition of $\mathrm{Seed}_{f,H}^-(\tau,T)$, for $k \in \{1,2\}$ there exists $j_k \in \mathrm{comp}_{14\tau}(\mathsf{R}_{f,H}^-(\tau,T),T) \subseteq \mathsf{R}_{f,H}^-(\tau,T)$ such that, letting $i_k = \alpha(j_k) = e_f^{\mathrm{full}}(\tau,T,j_k)$, it holds $e_k = \beta(j_k) = \min(7\tau,e_f^{\mathrm{full}}(\tau,T,j_k) - b(\tau,T,j_k))$, $w_k = |\{i' \in [1 \dots n] : T^\infty[i' - 7\tau \dots i' + 7\tau) = T^\infty[i_k - 7\tau \dots i_k + 7\tau)\}|$, and $\ell_k = \min\{i' \in [1 \dots n] : T^\infty[i' - 7\tau \dots i' + 7\tau) = T^\infty[i_k - 7\tau \dots i_k + 7\tau)\}$. For $k \in \{1,2\}$, denote $X_k = T^\infty[i_k - 7\tau \dots i_k + 7\tau)$. Note that for $k \in \{1,2\}$, it holds $T^\infty[\ell_k - 7\tau \dots \ell_k + 7\tau) = X_k$. Thus, the assumption $\ell_1 = \ell_2$ implies $\ell_1 = \ell_2 = T^\infty[\ell_1 - \tau \dots \ell_1 + \tau \tau) = T^\infty[\ell_2 - \tau \dots \ell_2 + \tau \tau) = X_k$. This in turn implies that it holds $\ell_1 = \ell_2 = T^\infty[\ell_1 - \tau \dots \ell_1 + \tau \tau) = T^\infty[\ell_2 - \tau \dots \ell_2 + \tau \tau) = X_k$. This in turn implies that it holds $\ell_1 = \ell_2 = T^\infty[\ell_1 - \tau \dots \ell_1 + \tau \tau) = T^\infty[\ell_1 - \tau \dots \ell_2 + \tau \tau) = X_k$. This in turn implies that it holds $\ell_1 = \ell_2 = T^\infty[\ell_1 - \tau \dots \ell_1 + \tau \tau) = T^\infty[\ell_2 - \tau \dots \ell_2 + \tau \tau) = X_k$. Thus, the assumption $\ell_1 = \ell_2 = T^\infty[\ell_1 - \tau \dots \ell_1 + \tau \tau) = T^\infty[\ell_1 - \tau \dots \ell_2 + \tau \tau) = X_k$. This in turn implies that it holds $\ell_1 = T^\infty[\ell_1 - \tau \dots \ell_1 + \tau \tau) = T^\infty[\ell_2 - \tau \dots \ell_2 + \tau \tau) = X_k$. This in turn implies that it holds $\ell_1 = T^\infty[\ell_1 - \tau \dots \ell_1 + \tau \tau) = T^\infty[\ell_1 - \tau \dots \ell_2 + \tau \tau) = X_k$. This in turn implies that it holds $\ell_1 = T^\infty[\ell_1 - \tau \dots \ell_1 + \tau \tau) = T^\infty[\ell_1 - \tau \dots \ell_2 + \tau \tau) = X_k$. This is turn implies that it holds $\ell_1 = T^\infty[\ell_1 - \tau \dots \ell_1 + \tau \tau) = T^\infty[\ell_1 - \tau \dots \ell_2 + \tau \tau) = X_k$. This is turn implies that it holds $\ell_1 = T^\infty[\ell_1 - \tau \dots \ell_1 + \tau \tau) = T^\infty[\ell_1 - \tau \dots \ell_2 + \tau \tau) = X_k$. This is turn implies that it holds $\ell_1 = T^\infty[\ell_1 - \tau \dots \ell_2 + \tau \tau] = T^\infty[\ell_1 - \tau \dots \ell_2 + \tau \tau]$. This is turn impl

A Necklace-Consistent Function

Definition 8.64. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$. Let $(r_i)_{i \in [1 ... q]}$ be a sequence containing all elements of $\mathsf{R}'(\tau, T)$ in sorted order, i.e, for any $i, i' \in [1 ... q], i < i'$ implies $r_i < r_{i'}$. Denote

$$T' = \prod_{i=1,2,\dots,q} T[r_i \dots r_i + 2\tau - 1)T[n] \in \Sigma^{2\tau q},$$

$$Q = \bigcup_{i \in \mathbb{R}(\tau,T)} \{T[i+\delta \dots i+\delta + \operatorname{per}(T[i\dots i+3\tau-1))) : \delta \in [0\dots\tau)\} \subseteq \Sigma^{\leq \tau/3},$$

and let

- $g: \mathbb{Z} \to \mathbb{Z}$ be defined by $g(i) = 1 + 2\tau \lfloor \frac{i-1}{2\tau} \rfloor$, $h: \mathcal{Q} \to \Sigma^+$ be defined by $h(X) = T'[g(s) \dots g(s) + |X|)$, where $s = \min \operatorname{Occ}(X^{\infty}[1 \dots \tau], T')$,
- $h': \Sigma^+ \to \Sigma^+$ be defined by $h'(X) = \min\{\operatorname{rot}^i(X) : i \in [0..|X|)\}.$

Then, by $f_{\tau,T}: \Sigma^+ \to \Sigma^+$ we define the function such that for every $X \in \Sigma^+$,

$$f_{\tau,T}(X) = \begin{cases} h(X) & \text{if } X \in \mathcal{Q}, \\ h'(X) & \text{otherwise.} \end{cases}$$

Lemma 8.65. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$. Then, $f_{\tau,T}$ (Definition 8.64) is well-defined.

Proof. Recall that $R'(\tau,T) \subseteq R(\tau,T) \subseteq [1..n-3\tau+2]$. Thus, for $i \in [1..q]$, it holds $r_i+2\tau-1 \le n-\tau+1 \le n$. Hence, $T[r_i cdots r_i + 2\tau - 1)$ in the formula for T' is well-defined.

Next, we show that for every $X \in \mathcal{Q}$, the value $\min \operatorname{Occ}(X^{\infty}[1..\tau], T')$ is well-defined, i.e., that $\operatorname{Occ}(X^{\infty}[1..\tau],T')\neq\emptyset$. Let $j\in\mathsf{R}(\tau,T)$ and $\delta\in[0..\tau)$ be such that $X=T[j+\delta..j+\delta+p)$, where $p = per(T[j ... j + 3\tau - 1))$. Observe, that p being a period of $T[j ... j + 3\tau - 1)$ implies that $T[j + \delta ... j + 3\tau - 1)$ is a prefix of X^{∞} . Thus, by $\delta < \tau$, it holds $X^{\infty}[1..\tau] = T[j+\delta..j+\delta+\tau)$. Note also that $p \leq \frac{1}{3}\tau$. Let $j' \in [1..j]$ be the smallest integer such that $[j'..j] \subseteq \mathsf{R}(\tau,T)$. Then, $j' \in \mathsf{R}'(\tau,T)$. Consider the substring $T[j' ... e(\tau, T, j'))$. On the one hand, by $[j' ... j] \subseteq R(\tau, T)$ and Lemma 8.49(1), it holds $e(\tau, T, j') \ge j + 3\tau - 1$, and hence $j' \leq j + \delta < j + \delta + \tau < j + 3\tau - 1 \leq e(\tau, T, j')$. Thus, $X^{\infty}[1..\tau]$ is a substring of $T[j'..e(\tau, T, j'))$. On the other hand, by Lemma 8.48, $\operatorname{per}(T[j'..j'+3\tau-1]) = \operatorname{per}(T[j..j+3\tau-1]) = p$. By definition of $e(\tau, T, j')$, $T[j' ... e(\tau, T, j'))$ thus has period p. Consequently, there exists $\delta' \in [0...p)$ such that $T[j'+\delta'\ldots j'+\delta'+\tau)=X^{\infty}[1\ldots\tau].$ Note that by $\delta'< p\leq \frac{1}{3}\tau\leq \tau,$ we then have $j'+\delta'+\tau\leq j'+2\tau-1.$ Consequently, letting $i \in [1..q]$ be such that $r_i = j'$, we have that $X^{\infty}[1..\tau]$ is a substring of $T[r_i..r_i+2\tau-1)$, and hence $Occ(X^{\infty}[1..\tau], T') \neq \emptyset$.

Lemma 8.66. Let $\tau \in [1... | \frac{n}{2}]$. Then, $f_{\tau,T}$ (Definition 8.64) is necklace-consistent.

Proof. First, we show that for any $X \in \Sigma^+$, the strings $f_{\tau,T}(X)$ and X are cyclically equivalent. Consider two cases:

- 1. Let us first assume $X \in \mathcal{Q}$. Denote $s = \min \operatorname{Occ}(X^{\infty}[1..\tau], T')$ and $i = \lceil \frac{s}{2\tau} \rceil$. We prove the following three properties.
 - (a) First, we show that $[s ... s + \tau) \subseteq [1 + 2\tau(i-1) ... 2\tau i)$, i.e., $X^{\infty}[1 ... \tau]$ is a substring of $T[r_i ... r_i + 2\tau 1)$. By definition of \mathcal{Q} , there exists $j \in \mathbb{R}(\tau, T)$ and $\delta \in [0...\tau)$ such that $X = T[j + \delta...j + \delta + p)$, where $p = \operatorname{per}(T[j ... j + 3\tau - 1)) \leq \frac{1}{3}\tau$. By $j \in \mathsf{R}(\tau, T)$, we have $j \in [1 ... n - 3\tau + 2]$. Therefore, $j + \delta + p \le j + 2\tau - 1 \le n - \tau + 1 \le n$. Consequently, T[n] does not occur in X, and hence by definition of T', we have $[s ... s + \tau) \subseteq [1 + 2\tau(i-1)... 2\tau i)$. Note that for such s, we have $g(s) = 1 + 2\tau \lfloor \frac{s-1}{2\tau} \rfloor = 1 + 2\tau (i-1).$
 - (b) Second, we show that $per(T[r_i ... r_i + 3\tau 1)) = |X|$. Denote $p' = per(T[r_i ... r_i + 3\tau 1))$ and let $\delta' \in [0..\tau)$ be such that $T[r_i + \delta'..r_i + \delta' + \tau] = X^{\infty}[1..\tau]$. Then, $T[r_i + \delta'..r_i + \delta' + \tau]$ has periods |X| and p'. By $|X|, p' \leq \frac{1}{3}\tau$ and the weak periodicity lemma [FW65], $\gcd(|X|, p')$ is therefore a period of $T[r_i + \delta' ... r_i + \delta' + \tau)$. Since every length-p' substring of $T[r_i ... r_i + 3\tau - 1)$

- is primitive, we thus have gcd(|X|, p') = p'. Consequently, $p' \mid |X|$. On the other hand, X being a substring of $T[j ... j + 3\tau 1)$ and $per(T[j ... j + 3\tau 1)) = |X|$ imply that X is primitive. Thus, p' = |X|.
- (c) Third, we show that $T[r_i + t ... r_i + t + |X|] = X$ holds for some $t \in [0...|X|]$. Since X is a substring of $T[r_i ... r_i + 3\tau 1)$ and $\operatorname{per}(T[r_i ... r_i + 3\tau 1)) = |X|$, we can write $T[r_i ... r_i + 3\tau 1) = X'X^kX''$, where $k \geq 1$ and X' (resp. X'') is a proper suffix (resp. prefix) of X. Letting t = |X'|, we obtain the claim. Note that we then also have $T[r_i ... r_i + |X|] = \operatorname{rot}^t(X)$.

Combining the above properties, we thus obtain

$$f_{\tau,T}(X) = h(X)$$

$$= T'[g(s) ... g(s) + |X|)$$

$$= T'[1 + 2\tau(i-1) ... 1 + 2\tau(i-1) + |X|)$$

$$= T[r_i ... r_i + |X|)$$

$$= \cot^t(X).$$

In other words, $f_{\tau,T}(X)$ and X are cyclically equivalent.

2. Let us now assume $X \notin \mathcal{Q}$. Then, $f_{\tau,T}(X) = \min\{\operatorname{rot}^i(X) : i \in [0..|X|)\}$. Since every element of $\{\operatorname{rot}^i(X) : i \in [0..|X|)\}$ is cyclically equivalent with X by definition, the claim follows.

We now show the second condition in Definition 8.44 for $f_{\tau,T}$. Let $X, X' \in \Sigma^+$ and assume that X and X' are cyclically equivalent. Consider two cases:

1. Let $X \in \mathcal{Q}$. Denote $s = \min \operatorname{Occ}(X^{\infty}[1 \dots \tau], T')$ and $i = \lceil \frac{s}{2\tau} \rceil$. By the three properties of $T[r_i \dots r_i + 3\tau - 1)$ proved above, it holds $\{T[r_i + t \dots r_i + t + |X|) : t \in [0 \dots |X|)\} = \{\operatorname{rot}^i(X) : i \in [0 \dots |X|)\}$. Since X' is in the latter set, there exists $\delta' \in [0 \dots |X'|)$ such that $T[r_i + \delta' \dots r_i + \delta' + |X'|) = X'$. This implies $X' \in \mathcal{Q}$. Moreover, since $\operatorname{per}(T[r_i \dots r_i + 3\tau - 1)) = |X| = |X'|$, we then have $T[r_i + \delta' \dots r_i + \delta' + \tau) = T'[1 + 2\tau(i-1) + \delta' \dots 1 + 2\tau(i-1) + \delta' + \tau) = X'^{\infty}[1 \dots \tau]$. Consequently, letting $s' = \min \operatorname{Occ}(X'^{\infty}[1 \dots \tau], T')$, we have $s' < 1 + 2\tau i$ and hence $g(s) = 1 + 2\tau(i-1) \ge 1 + 2\tau \lfloor \frac{s'-1}{2\tau} \rfloor = g(s')$. By $X' \in \mathcal{Q}$ and the symmetry of our argument, we analogously obtain $g(s) \le g(s')$. Thus, g(s) = g(s') and hence

$$\begin{split} f_{\tau,T}(X) &= h(X) \\ &= T'[g(s) \dots g(s) + |X|) \\ &= T'[g(s') \dots g(s') + |X'|) \\ &= h(X') \\ &= f_{\tau,T}(X'). \end{split}$$

2. Let us now assume $X \notin \mathcal{Q}$. By the above, we then have $X' \notin \mathcal{Q}$ and thus $f_{\tau,T}(X) = h'(X) = \min\{\operatorname{rot}^i(X) : i \in [0..|X|)\} = \min\{\operatorname{rot}^i(X') : i \in [0..|X'|)\} = h'(X') = f_{\tau,T}(X')$.

Next, we first describe an alternative construction of the necklace-consistent function from Definition 8.64, that instead of the whole set $R(\tau,T)$ represented using $R'(\tau,T)$, utilizes the set $comp_k(R(\tau,T),T)$. This will be used during the construction algorithms. After introducing it in Definition 8.67, we prove that it is well-defined (Lemma 8.68) and indeed equal to $f_{\tau,T}$ (Lemma 8.69) for every $k \geq 3\tau - 1$.

Definition 8.67. Let $\tau \in [1..\lfloor \frac{n}{2} \rfloor]$ and $k \geq 3\tau - 1$. Let \mathcal{Q} , g, and h' be as in Definition 8.64. Let $(a_i, t_i)_{i \in [1..q']} = \mathcal{I}(\text{comp}_k(\mathsf{R}(\tau, T), T))$. Denote

$$T'_{\text{comp}} = \prod_{i=1,2,...,q'} T[a_i ... a_i + 2\tau - 1)T[n] \in \Sigma^{2\tau q'}.$$

Let $h_{\text{comp}}: \mathcal{Q} \to \Sigma^+$ be a function such that for every $X \in \mathcal{Q}$, it holds $h_{\text{comp}}(X) = T'_{\text{comp}}[g(s) ... g(s) + |X|)$, where $s = \min \text{Occ}(X^{\infty}[1...\tau], T'_{\text{comp}})$. Finally, let $f_{k,\tau,T}: \Sigma^+ \to \Sigma^+$ be a function such that for every $X \in \Sigma^+$,

$$f_{k,\tau,T}(X) = \begin{cases} h_{\text{comp}}(X) & \text{if } X \in \mathcal{Q}, \\ h'(X) & \text{otherwise.} \end{cases}$$

Lemma 8.68. Let $\tau \in [1, \lfloor \frac{n}{2} \rfloor]$ and $k \geq 3\tau - 1$. Then, $f_{k,\tau,T}$ (Definition 8.67) is well-defined.

Proof. For every $i \in [1..q']$, we have $a_i \in \text{comp}_k(\mathsf{R}(\tau,T),T)$. On the other hand, by Definition 4.10, we have $\text{comp}_k(\mathsf{R}(\tau,T),T) = \mathsf{R}(\tau,T) \cap \mathsf{C}(k,T) \subseteq \mathsf{R}(\tau,T) \subseteq [1..n-3\tau+2]$. Thus, for every $i \in [1..q']$, it holds $a_i + 2\tau - 1 \le n - \tau + 1 \le n$, and hence $T[a_i ... a_i + 2\tau - 1)$ in the formula for T'_{comp} is well-defined.

We now show that for every $X \in \mathcal{Q}$, it holds $\operatorname{Occ}(X^{\infty}[1 \dots \tau], T'_{\operatorname{comp}}) \neq \emptyset$. By definition of \mathcal{Q} , there exists $j \in \mathsf{R}(\tau,T)$ and $\delta \in [0 \dots \tau)$ such that $X = T[j+\delta \dots j+\delta+p)$, where $p = \operatorname{per}(T[j \dots j+3\tau-1))$. In Lemma 8.65, we proved that we then have $T[j+\delta \dots j+\delta+\tau) = X^{\infty}[1 \dots \tau]$. Let $j' = \min \operatorname{Occ}_k(j,T)$. Note that $j' \leq j \leq n-3\tau+2$ (where the latter follows by $j \in \mathsf{R}(\tau,T) \subseteq [1 \dots n-3\tau+2]$). Thus, by $j' \in \operatorname{Occ}_k(j,T)$ and $k \geq 3\tau-1$ we have $T[j' \dots j'+3\tau-1) = T[j \dots j+3\tau-1)$. In particular, $T[j'+\delta \dots j'+\delta+\tau) = X^{\infty}[1 \dots \tau]$. Note now that by Lemma 8.16, we have $j' \in \operatorname{comp}_k(\mathsf{R}(\tau,T),T)$. Let $j'' \in [1 \dots j']$ be the smallest integer such that $[j'' \dots j'] \subseteq \operatorname{comp}_k(\mathsf{R}(\tau,T),T)$. In particular, we then have $[j'' \dots j'] \subseteq \mathsf{R}(\tau,T)$. In the proof of Lemma 8.65 we observed that this, combined with $X^{\infty}[1 \dots \tau]$ being a substring of $T[j' \dots j'+3\tau-1)$, implies that $T[j'' \dots e(\tau,T,j''))$ has period |X| and that $X^{\infty}[1 \dots \tau]$ is a substring of $T[j'' \dots e(\tau,T,j''))$. Thus, there exists $\delta' \in [0 \dots |X|)$ such that $T[j''+\delta' \dots j''+\delta'+\tau) = X^{\infty}[1 \dots \tau]$. Letting $i \in [1 \dots q']$ be such that $a_i = j''$, we obtain by $\delta' < \tau$, that $X^{\infty}[1 \dots \tau]$ is a substring of $T[a_i \dots a_i + 2\tau - 1)$, and therefore $\operatorname{Occ}(X^{\infty}[1 \dots \tau], T'_{\operatorname{comp}}) \neq \emptyset$. \square

Lemma 8.69. Let $\tau \in [1... | \frac{n}{2}]$ and $k \geq 3\tau - 1$. Then, it holds $f_{k,\tau,T} = f_{\tau,T}$ (Definitions 8.64 and 8.67).

Proof. By definition, we have $f_{k,\tau,T}(X) = f_{\tau,T}(X)$ for every $X \in \Sigma^+ \setminus \mathcal{Q}$. Thus, it remains to show that $h_{\text{comp}} = h$ (where h is as in Definition 8.64), i.e., that for every $X \in \mathcal{Q}$, it holds $T'_{\text{comp}}[g(s_{\text{comp}}) \dots g(s_{\text{comp}}) + |X|) = T'[g(s) \dots g(s) + |X|)$, where $s_{\text{comp}} = \min \operatorname{Occ}(X^{\infty}[1 \dots \tau], T'_{\text{comp}})$, $s = \min \operatorname{Occ}(X^{\infty}[1 \dots \tau], T')$, and T' is as in Definition 8.64. Let $(r_i)_{i \in [1 \dots q]}$ and $(a_i, t_i)_{i \in [1 \dots q']}$ be as in Definition 8.64 and Definition 8.67, respectively. Denote $i = \lceil \frac{s}{2\tau} \rceil$ and $k = \lceil \frac{s_{\text{comp}}}{2\tau} \rceil$.

- 1. First, we prove that it holds $a_k \leq r_i$. To achieve this, we will show that $r_i \in C(k,T)$. Suppose that $r_i \notin C(k,T)$. Then, $\min Occ_k(r_i,T) < r_i$ (see Lemma 4.5), i.e., there exists $j \in [1...r_i)$ such that $T^{\infty}[j ... j + k) = T^{\infty}[r_i ... r_i + k)$. Note that $j < r_i \le n - 3\tau + 2$ (where the latter follows by $r_i \in \mathsf{R}(\tau,T) \subseteq [1\ldots n-3\tau+2]$). Thus, $T[j\ldots j+3\tau-1)=T[r_i\ldots r_i+3\tau-1)$. In the proof of Lemma 8.66, we showed that $X^{\infty}[1..\tau]$ is a substring of $T[r_i..r_i+2\tau-1)$ and $\operatorname{per}(T[r_i..r_i+3\tau-1))=|X|$. Consequently, $\operatorname{per}(T[j\mathinner{\ldotp\ldotp} j+3\tau-1))=\operatorname{per}(T[r_i\mathinner{\ldotp\ldotp} r_i+3\tau-1))=|X|\leq \frac{1}{3}\tau$ and thus $j\in\mathsf{R}(\tau,T)$. Let $j' \in [1 ... j]$ be the smallest integer such that $[j' ... j] \subseteq \mathsf{R}(\tau, T)$. By Lemma 8.48, $\mathsf{per}(T[j' ... j' + 3\tau - 1)) =$ $\operatorname{per}(T[j ... j + 3\tau - 1)) = |X|$. Thus, by definition of $e(\tau, T, j')$, the string $T[j' ... e(\tau, T, j'))$ has period |X|. Since $X^{\infty}[1..\tau]$ is the substring of $T[j'..e(\tau,T,j'))$, we therefore obtain $\delta \in [0..|X|)$ satisfying $T[j'+\delta ...j'+\delta +\tau)=X^{\infty}[1...\tau]$, i.e., $X^{\infty}[1...\tau]$ is a substring of $T[j'...j'+2\tau-1)$. Equivalently, letting $t \in [1..q]$ be such that $r_t = j'$ (such t exists since $j' \in \mathsf{R}'(\tau,T)$), $X^{\infty}[1..\tau]$ is a substring of $T'[1+2\tau(t-1)..2\tau t)$. Since $j',r_i \in R'(\tau,T)$ and $j' \leq j < r_i$, we have t < i. Therefore, $\min \operatorname{Occ}(X^{\infty}[1..\tau], T') < 1 + 2\tau t \le 1 + 2\tau (i-1) \le s$, which contradicts the definition of s. We have thus proved $r_i \in C(k,T)$. Combined with $r_i \in R'(\tau,T)$, this implies that there exists $h \in [1..q']$ such that $a_h = r_i$. Since $X^{\infty}[1..\tau]$ is a substring of $T[r_i..r_i + 2\tau - 1)$, we thus obtain that the leftmost occurrence of $X^{\infty}[1..\tau]$ in T'_{comp} occurs either inside or to the left of the block $T[a_h..a_h + 2\tau - 1)T[n]$. More precisely, $s_{\text{comp}} < 1 + 2\tau h$. This implies $k = \lceil \frac{s_{\text{comp}}}{2\tau} \rceil \le h$ and hence $a_k \le a_h = r_i$.
- 2. Second, we prove that it holds $a_k \geq r_i$. Suppose that $a_k < r_i$. From the definition of k, it follows that $X^{\infty}[1..\tau]$ occurs inside $T'_{\text{comp}}[1+2\tau(k-1)..2\tau k) = T[a_k..a_k+2\tau-1)$. Recall that $a_k \in \text{comp}_k(\mathsf{R}(\tau,T),T) \subseteq \mathsf{R}(\tau,T)$. Let $j \in [1..a_k]$ be the smallest integer such that $[j..a_k] \subseteq \mathsf{R}(\tau,T)$. Consider the substring $T[j..e(\tau,T,j))$. In the proof of Lemma 8.65, we showed that then $T[j..e(\tau,T,j))$ has period |X| and $X^{\infty}[1..\tau]$ is a substring of $T[j..e(\tau,T,j))$. Thus, there exists $\delta \in [0..|X|)$ such that $T[j+\delta..j+\delta+\tau) = X^{\infty}[1..\tau]$. Note also that $j \in \mathsf{R}'(\tau,T)$. Letting $i' \in [1..q]$ be such that $r_{i'}=j$, we have thus obtained an occurrence of $X^{\infty}[1..\tau]$ in $T[r_{i'}..r_{i'}+2\tau-1)$. Note that since

 $j \leq a_k < r_i$, we have i' < i. By $T[r_{i'} ... r_{i'} + 2\tau - 1) = T'[1 + 2\tau(i' - 1)... 2\tau i')$, we therefore have $\min \operatorname{Occ}(X^{\infty}[1..\tau], T') < 2\tau i' < 1 + 2\tau (i-1) \le s$, which contradicts the definition of s.

By the above, we thus have $a_k = r_i$. Consequently,

$$\begin{split} h_{\text{comp}}(X) &= T'_{\text{comp}}[g(s_{\text{comp}}) \cdot \cdot g(s_{\text{comp}}) + |X|) \\ &= T'_{\text{comp}}[1 + 2\tau(k-1) \cdot \cdot 1 + 2\tau(k-1) + |X|) \\ &= T[a_k \cdot \cdot \cdot a_k + |X|) \\ &= T[r_i \cdot \cdot \cdot r_i + |X|) \\ &= T'[1 + 2\tau(i-1) \cdot \cdot \cdot 1 + 2\tau(i-1) + |X|) \\ &= T'[g(s) \cdot \cdot \cdot g(s) + |X|) \\ &= h(X). \end{split}$$

8.4.3 Decomposition Lemmas

Definition 8.70. Let f be a necklace-consistent function. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. For every τ -periodic pattern $P \in \Sigma^+$, we define:

```
\begin{split} & \operatorname{Pos}^{\operatorname{low}-}_{f,\ell}(P,T) := \{j' \in \mathsf{R}^-_{f,s,H}(\tau,T) : \exp_f(\tau,T,j') = k_1 \text{ and } (T[j'\mathinner{.\,.} n] \succeq P \text{ or } \operatorname{lcp}(P,T[j'\mathinner{.\,.} n]) \geq \ell)\}, \\ & \operatorname{Pos}^{\operatorname{mid}-}_{f,\ell}(P,T) := \{j' \in \mathsf{R}^-_{f,s,H}(\tau,T) : \exp_f(\tau,T,j') \in (k_1\mathinner{.\,.} k_2]\}, \\ & \operatorname{Pos}^{\operatorname{high}-}_{f,\ell}(P,T) := \{j' \in \mathsf{R}^-_{f,s,H}(\tau,T) : \exp_f(\tau,T,j') = k_2 \text{ and } (T[j'\mathinner{.\,.} n] \succeq P \text{ or } \operatorname{lcp}(P,T[j'\mathinner{.\,.} n]) \geq 2\ell)\}, \end{split}
```

where $s = \text{head}_f(\tau, P)$, $H = \text{root}_f(\tau, P)$, $k_1 = \exp^{\text{cut}}_f(\tau, P, \ell)$, and $k_2 = \exp^{\text{cut}}_f(\tau, P, 2\ell)$. We denote $\delta^{\text{low}-}_{f,\ell}(P,T) := |\text{Pos}^{\text{low}-}_{f,\ell}(P,T)|$, $\delta^{\text{mid}-}_{f,\ell}(P,T) := |\text{Pos}^{\text{mid}-}_{f,\ell}(P,T)|$, and $\delta^{\text{high}-}_{f,\ell}(P,T) := |\text{Pos}^{\text{high}-}_{f,\ell}(P,T)|$. Next, by $\text{Pos}^{\text{low}+}_{f,\ell}(P,T)$, $\text{Pos}^{\text{mid}+}_{f,\ell}(P,T)$, and $\text{Pos}^{\text{high}+}_{f,\ell}(P,T)$ we denote the symmetric versions of the above sets (i.e., with $\mathsf{R}^-_{f,s,H}(\tau,T)$ replaced by $\mathsf{R}^+_{f,s,H}(\tau,T)$, and the condition $T[j':n] \succeq P$ replaced with $T[j':n] \succeq_{\text{inv}} P$). Finally, we denote $\text{Pos}^{\text{low}}_{f,\ell}(P,T) := \text{Pos}^{\text{low}-}_{f,\ell}(P,T) \cup \text{Pos}^{\text{low}+}_{f,\ell}(P,T)$, $\text{Pos}^{\text{mid}}_{f,\ell}(P,T) := \text{Pos}^{\text{mid}-}_{f,\ell}(P,T) \cup \text{Pos}^{\text{high}+}_{f,\ell}(P,T)$.

Remark 8.71. Observe that, similarly as when characterizing $\operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T)$ (Lemma 8.4; see also Remark 8.5), it is not correct to simply change \succeq to \preceq when defining the symmetric versions of sets $\operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T)$, $\operatorname{Pos}_{f,\ell}^{\operatorname{mid}-}(P,T)$, and $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$. Instead, we replace \succeq in Definition 8.70 with $\succeq_{\operatorname{inv}}$, where the inverted lexicographic order is as in Definition 8.2. Observe that then $\mathsf{R}^+_{f,H}(\tau,T)$ from the standard order becomes $\mathsf{R}_{f,H}^-(\tau,T)$ in the inverted order. Thus, the combinatorial results and query algorithms concerning $\mathsf{Pos}^{\mathsf{low}+}_{f,\ell}(P,T)$, $\mathsf{Pos}^{\mathsf{mid}+}_{f,\ell}(P,T)$, and $\mathsf{Pos}^{\mathsf{high}+}_{f,\ell}(P,T)$ are identical as for $\mathsf{Pos}^{\mathsf{low}-}_{f,\ell}(P,T)$, $\mathsf{Pos}^{\mathsf{mid}-}_{f,\ell}(P,T)$, and $\mathsf{Pos}^{\mathsf{high}-}_{f,\ell}(P,T)$ (with the inverted order). This is why the structure from Section 8.4.4 contains two parts: one for the standard lexicographic order \leq , and one for the inverted lexicographic order \leq inv.

Lemma 8.72. Let f be a necklace-consistent function. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let $P \in \Sigma^+$ be a τ -periodic pattern and let m = |P|. Then:

- 1. Let $P' = P[1 ... \min(m, \ell)]$. Then, P' is τ -periodic and $\operatorname{Pos}^{\operatorname{low}-}_{f, \ell}(P, T) = \operatorname{Pos}^{\operatorname{low}-}_{f, \ell}(P', T)$, 2. Let $P' = P[1 ... \min(m, 2\ell)]$. Then, P' is τ -periodic and $\operatorname{Pos}^{\operatorname{mid}-}_{f, \ell}(P, T) = \operatorname{Pos}^{\operatorname{mid}-}_{f, \ell}(P', T)$, 3. Let $P' = P[1 ... \min(m, 2\ell)]$. Then, P' is τ -periodic and $\operatorname{Pos}^{\operatorname{high}-}_{f, \ell}(P, T) = \operatorname{Pos}^{\operatorname{high}-}_{f, \ell}(P', T)$.

 $Proof. \ \ \text{Denote} \ \ H = \text{root}_f(\tau, P), \ s = \text{head}_f(\tau, P), \ p = |H|, \ k_1 = \exp_f^{\text{cut}}(\tau, P, \ell), \ \text{and} \ \ k_2 = \exp_f^{\text{cut}}(\tau, P, 2\ell).$

- 1. By $m \geq 3\tau 1$ and $\ell \geq 3\tau 1$, we obtain $|P'| \geq 3\tau 1$. Thus, it follows by Lemma 8.45 that P'is τ -periodic. Moreover, by the same result, we also have root $f(\tau, P') = H$ and head $f(\tau, P') = s$. Denote $k_1' = \exp_f^{\text{cut}}(\tau, P', \ell)$. Note that by Lemma 8.47(1), it follows that $k_1' = \exp_f^{\text{cut}}(\tau, P', \ell) = \exp_f^{\text{cut}}(\tau, P, \ell) = k_1$. Given the above, we thus prove the two inclusions as follows:
 - Let $j \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T)$. Then, it holds $j \in \mathsf{R}_{f,s,H}^-(\tau,T)$, $\exp_f(\tau,T,j) = k_1$, and $T[j\mathinner{\ldotp\ldotp} n] \succeq P$ or $\operatorname{lcp}(P,T[j\mathinner{\ldotp\ldotp} n]) \geq \ell$. By the above, to show $j \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P',T)$, it thus remains to prove that

- either $T[j ... n] \succeq P'$ or $lcp(P', T[j ... n]) \geq \ell$. By Lemma 8.8(1), the assumptions $T[j ... n] \succeq P$ or $\operatorname{lcp}(P, T[j \dots n]) \geq \ell \text{ imply } T[j \dots n] \succeq P[1 \dots \min(m, \ell)] = P'. \text{ We have thus proved } j \in \operatorname{Pos}_{f, \ell}^{\operatorname{low}^-}(P', T).$
- Let $j' \in \operatorname{Pos}_{f,\ell}^{low-}(P',T)$. Then, it holds $j \in \mathsf{R}_{f,s,H}^{-}(\tau,T)$, $\exp_f(\tau,T,j) = k'_1$, and $T[j\mathinner{\ldotp\ldotp} n] \succeq P'$ or $\operatorname{lcp}(P',T[j\mathinner{\ldotp\ldotp} n]) \geq \ell$. By the above, to show $j \in \operatorname{Pos}_{f,\ell}^{low-}(P,T)$, it thus remains to prove that either $T[j ... n] \succeq P$ or $lcp(P, T[j ... n]) \ge \ell$. Let us consider two cases. If $T[j ... n] \succeq P' = P[1 ... min(m, \ell)]$, then the claim follows by Lemma 8.8(1). Otherwise (i.e., if $lcp(P', T[j ... n]) \ge \ell$), then $|P'| = \ell$, and hence $P' = P[1 ... \ell]$. This immediately implies $lcp(P, T[j ... n]) \ge \ell$. We have thus proved $j \in Pos_{f,\ell}^{low-}(P,T)$.
- 2. By $m \geq 3\tau 1$ and $2\ell \geq 3\tau 1$, we obtain $|P'| \geq 3\tau 1$. Thus, it follows by Lemma 8.45 that P'is τ -periodic. Moreover, by the same result, we also have $\operatorname{root}_f(\tau, P') = H$ and $\operatorname{head}_f(\tau, P') = s$. Denote $k_1' = \exp_f^{\text{cut}}(\tau, P', \ell)$ and $k_2' = \exp_f^{\text{cut}}(\tau, P', 2\ell)$. By Lemma 8.47(1), it follows that $k_1' = \exp_f^{\text{cut}}(\tau, P', \ell) = \exp_f^{\text{cut}}(\tau, P, \ell) = k_1$ and $k_2' = \exp_f^{\text{cut}}(\tau, P', 2\ell) = \exp_f^{\text{cut}}(\tau, P, 2\ell) = k_2$. We thus have $\operatorname{Pos}_{f,\ell}^{\operatorname{mid}}(P, T) = \operatorname{Pos}_{f,\ell}^{\operatorname{mid}}(P, T)$ Pos $_{f,\ell}^{\text{mid}-}(P',T)$.

 3. The proof follows analogously as in Lemma 8.72(1).

Remark 8.73. Note that the following lemma holds even when P does not occur in T (and we indeed use it in that case). Note also that it is important (Proposition 8.140) that the lemma below holds not only for patterns P satisfying $e(\tau, P) \leq |P|$ but also for the case $e(\tau, P) = |P| + 1$ (we have $type(\tau, P) = -1$ for such patterns).

Lemma 8.74. Let f be a necklace-consistent function. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let $P \in \Sigma^+$ be a τ -periodic pattern satisfying type $(\tau, P) = -1$. Then,

$$\delta_{\ell}^{\mathrm{beg}}(P,T) = \delta_{f,\ell}^{\mathrm{low-}}(P,T) + \delta_{f,\ell}^{\mathrm{mid-}}(P,T) - \delta_{f,\ell}^{\mathrm{high-}}(P,T).$$

 $\begin{array}{l} \textit{Proof.} \ \textit{First, observe that by definition it holds} \ \textit{Pos}^{\textit{beg}}_{\ell}(P,T) \cap \textit{Pos}^{\textit{high-}}_{f,\ell}(P,T) = \emptyset \ \text{and} \ \textit{Pos}^{\textit{low-}}_{f,\ell}(P,T) \cap \textit{Pos}^{\textit{high-}}_{f,\ell}(P,T) = \emptyset. \ \text{We will prove that} \ \textit{Pos}^{\textit{beg}}_{\ell}(P,T) \cup \textit{Pos}^{\textit{high-}}_{f,\ell}(P,T) = \textit{Pos}^{\textit{low-}}_{f,\ell}(P,T) \cup \textit{Pos}^{\textit{mid-}}_{f,\ell}(P,T), \ \text{since} \end{array}$ then we have

$$\begin{split} \delta^{\text{low}-}_{f,\ell}(P,T) + \delta^{\text{mid}-}_{f,\ell}(P,T) &= |\text{Pos}^{\text{low}-}_{f,\ell}(P,T)| + |\text{Pos}^{\text{mid}-}_{f,\ell}(P,T)| \\ &= |\text{Pos}^{\text{low}-}_{f,\ell}(P,T) \cup \text{Pos}^{\text{mid}-}_{f,\ell}(P,T)| \\ &= |\text{Pos}^{\text{beg}}_{\ell}(P,T) \cup \text{Pos}^{\text{high}-}_{f,\ell}(P,T)| \\ &= |\text{Pos}^{\text{beg}}_{\ell}(P,T)| + |\text{Pos}^{\text{high}-}_{f,\ell}(P,T)| \\ &= \delta^{\text{beg}}_{\ell}(P,T) + \delta^{\text{high}-}_{f,\ell}(P,T) \end{split}$$

which implies the claim. It therefore remains to show $\operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T) \cup \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) = \operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T) \cup \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$. Denote $s = \operatorname{head}_f(\tau,P)$ and $H = \operatorname{root}_f(\tau,P)$. Observe that for every $j' \in \operatorname{Pos}_{\ell}^{\operatorname{low}-}(P,T)$, it holds $lcp(P, T[j'..n]) \ge \ell \ge 3\tau - 1$. Thus, by Lemma 8.50(1), we have $j' \in R_{f,s,H}(\tau,T)$. We furthermore have type $(\tau, T, j') = -1$, since otherwise by Lemma 8.51(1) and Lemma 8.51(2), we would have $P \prec T[j' ... n]$, which would contradict $j' \in \operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T)$. Thus, $\operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T) \subseteq \operatorname{R}^-_{f,s,H}(\tau,T)$.

First first show the inclusion $\operatorname{Pos}^{\operatorname{low}}_{f,\ell}(P,T) \cup \operatorname{Pos}^{\operatorname{mid}}_{f,\ell}(P,T) \subseteq \operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T) \cup \operatorname{Pos}^{\operatorname{high}}_{f,\ell}(P,T)$. Let $j' \in \operatorname{Pos}^{\operatorname{low}}_{f,\ell}(P,T)$. Note that $k_1 \leq k_2$. Let us thus consider two cases:

• Let us first assume $k_1 < k_2$. This is equivalent to $\min(\exp_f(\tau, P), \lfloor \frac{\ell - s}{|H|} \rfloor) < \min(\exp_f(\tau, P), \lfloor \frac{2\ell - s}{|H|} \rfloor)$, which implies $\lfloor \frac{\ell - s}{|H|} \rfloor < \exp_f(\tau, P)$. Therefore, $k_1 = \lfloor \frac{\ell - s}{|H|} \rfloor < \exp_f(\tau, P)$. Denote $t = e(\tau, T, j') - j'$ and $t' = e(\tau, P) - 1$. Consequently, by combining $\operatorname{type}(\tau, T, j') = -1$ and $t = e(\tau, T, j') - j' = -1$. $s+\exp_f(\tau,T,j')|H|+\operatorname{tail}_f(\tau,T,j')=s+k_1|H|+\operatorname{tail}_f(\tau,T,j')< s+(k_1+1)|H|\leq s+\exp_f(\tau,P)|H|\leq s+\exp_f(\tau,P)|H|$ $s + \exp_f(\tau, P)|H| + \operatorname{tail}_f(\tau, P) = e(\tau, P) - 1 = t'$, we obtain from Lemma 8.51(3) that $\mathring{T}[j' \dots n] \prec P$. By definition of $\operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T)$, this implies $\operatorname{lcp}(P,T[j'\ldots n]) \geq \ell$. On the other hand, by t < t' it follows from Lemma 8.51(1) that $lcp(P, T[j'..n]) = t = e(\tau, T, j') - j' = s + exp_f(\tau, T, j')|H| + tail_f(\tau, T, j') = s + exp_f(\tau, T, j')|H|$ $s + k_1|H| + \operatorname{tail}_f(\tau, T, j') = s + \lfloor \frac{\ell - s}{|H|} \rfloor |H| + \operatorname{tail}_f(\tau, T, j') \le \ell + \operatorname{tail}_f(\tau, T, j') < \ell + \tau < 2\ell$. We have thus proved that $T[j' \dots n] \prec P$ and $\operatorname{lcp}(P, T[j' \dots n]) \in [\ell \dots 2\ell)$, i.e., $j' \in \operatorname{Pos}_\ell^{\operatorname{beg}}(P, T)$.

• Let us now assume $k_1 = k_2$. By $j' \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T)$, this immediately implies $\exp_f(\tau,T,j') = k_2$. Consider two cases. If $T[j' \ldots n] \succeq P$ or $\operatorname{lcp}(P,T[j' \ldots n]) \geq 2\ell$, then we immediately obtain $j' \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P,T)$. Otherwise, we have $T[j' \ldots n] \prec P$ and $\operatorname{lcp}(P,T[j' \ldots n]) < 2\ell$. Combining this with $\operatorname{lcp}(P,T[j' \ldots n]) \geq \ell$ (following from $j' \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T)$), we thus have $j' \in \operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T)$.

We have thus proved $\operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T) \subseteq \operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T) \cup \operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T)$. To show $\operatorname{Pos}_{f,\ell}^{\operatorname{mid}}(P,T) \subseteq \operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T) \cup \operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T)$, consider $j' \in \operatorname{Pos}_{f,\ell}^{\operatorname{mid}}(P,T)$, and note that $\operatorname{Pos}_{f,\ell}^{\operatorname{mid}}(P,T) \neq \emptyset$ implies $k_1 < k_2$. As noted above, this implies $k_1 = \lfloor \frac{\ell - s}{|H|} \rfloor < \exp_f(\tau,P)$. Letting $t' = e(\tau,P) - 1$, we thus have $t' = s + \exp_f(\tau,P)|H| + \operatorname{tail}_f(\tau,P) \geq (\lfloor \frac{\ell - s}{|H|} \rfloor + 1)|H| \geq \ell$. On the other hand, $j' \in \operatorname{Pos}_{f,\ell}^{\operatorname{mid}}(P,T)$ implies $k_1 < \exp_f(\tau,T,j')$. Letting $t = e(\tau,T,j') - j'$, we thus have $t = s + \exp_f(\tau,T,j')|H| + \operatorname{tail}_f(\tau,T,j') \geq (k_1 + 1)|H| = (\lfloor \frac{\ell - s}{|H|} \rfloor + 1)|H| \geq \ell$. Thus, by Lemma 8.51, we obtain $\operatorname{lcp}(P,T[j' \dots n]) \geq \min(t,t') \geq \ell$. Recall now that $\exp_f(\tau,T,j') \in (k_1 \dots k_2]$, and consider two cases:

- First, assume $\exp_f(\tau, T, j') = k_2$. Then, either at least one of the following conditions: (1) $T[j' ... n] \ge P$, (2) $\operatorname{lcp}(P, T[j' ... n]) \ge 2\ell$, is true, or both are false (i.e., it holds $T[j' ... n] \prec P$ and $\operatorname{lcp}(P, T[j' ... n]) < 2\ell$). If the first possibility holds, then we immediately obtain $j' \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P, T)$. Otherwise, by combining with $\operatorname{lcp}(P, T[j' ... n]) \ge \ell$, we have $j' \in \operatorname{Pos}_{\ell}^{\operatorname{beg}}(P, T)$.
- Next, assume $\exp_f(\tau, T, j') \in (k_1 \dots k_2)$. In particular, $\exp_f(\tau, T, j') < k_2 = \min(\exp_f(\tau, P), \lfloor \frac{2\ell s}{|H|} \rfloor)$, i.e., $\exp_f(\tau, T, j') < \exp_f(\tau, P)$ and $\exp_f(\tau, T, j') < \lfloor \frac{2\ell s}{|H|} \rfloor$. The first inequality implies $\exp_f(\tau, T, j') j' = s + \exp_f(\tau, T, j') |H| + \tan|_f(\tau, T, j') < s + (\exp_f(\tau, T, j') + 1) |H| \le s + \exp_f(\tau, P) |H| \le s + \exp_f(\tau, P) |H| + \tan|_f(\tau, P) = e(\tau, P) 1$. By the second inequality, on the other hand, $\exp_f(\tau, T, j') j' = s + \exp_f(\tau, T, j') |H| + \tan|_f(\tau, T, j') < s + (\exp_f(\tau, T, j') + 1) |H| \le s + \lfloor \frac{2\ell s}{|H|} \rfloor |H| \le 2\ell$. By Lemma 8.51(1), we thus obtain $\log(P, T[j' \dots n]) = \min(e(\tau, T, j') j', e(\tau, P) 1) = e(\tau, T, j') j' < 2\ell$. Since above we proved $\log(P, T[j' \dots n]) \ge \ell$, we thus have $\log(P, T[j' \dots n]) \in [\ell \dots 2\ell)$. Lastly, it follows from $e(\tau, T, j') j' < e(\tau, P) 1$ by applying Lemma 8.51(3) that $T[j' \dots n] \prec P$. We have therefore proved $j' \in \operatorname{Pos}_{\ell}^{\operatorname{beg}}(P, T)$.

We have thus proved $\operatorname{Pos}^{\operatorname{mid}-}_{f,\ell}(P,T)\subseteq \operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T)\cup \operatorname{Pos}^{\operatorname{high}-}_{f,\ell}(P,T)$. Combining this with the above proof of $\operatorname{Pos}^{\operatorname{low}-}_{f,\ell}(P,T)\subseteq \operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T)\cup \operatorname{Pos}^{\operatorname{high}-}_{f,\ell}(P,T)$, we obtain $\operatorname{Pos}^{\operatorname{low}-}_{f,\ell}(P,T)\cup \operatorname{Pos}^{\operatorname{mid}-}_{f,\ell}(P,T)\subseteq \operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T)\cup \operatorname{Pos}^{\operatorname{high}-}_{f,\ell}(P,T)$.

We now prove the opposite inclusion, i.e., $\operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T) \cup \operatorname{Pos}^{\operatorname{high}-}_{f,\ell}(P,T) \subseteq \operatorname{Pos}^{\operatorname{low}-}_{f,\ell}(P,T) \cup \operatorname{Pos}^{\operatorname{mid}-}_{f,\ell}(P,T)$. Consider $j' \in \operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T)$. First, observe that we have $e(\tau,T,j')-j' \leq e(\tau,P)-1$, i.e., $e(\tau,T,j')-j' = \min(e(\tau,T,j')-j',e(\tau,P)-1)$ since otherwise, by Lemma 8.51(5) we would have $T[j'..n] \succeq P$, contradicting $j' \in \operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T)$. Second, note that by Lemma 8.51, we have $\min(e(\tau,T,j')-j',e(\tau,P)-1) \leq \operatorname{lcp}(P,T[j'..n])$. Third, by $j' \in \operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T)$, we have $\operatorname{lcp}(P,T[j'..n]) < 2\ell$. Combining these three inequalities, we obtain $e(\tau,T,j')-j'=\min(e(\tau,T,j')-j',e(\tau,P)-1) \leq \operatorname{lcp}(P,T[j'..n]) < 2\ell$, which implies $\exp_f(\tau,T,j')=\lfloor\frac{e(\tau,T,j')-j'-s}{|H|}\rfloor \leq \lfloor\frac{2\ell-s}{|H|}\rfloor$. On the other hand, from $e(\tau,T,j')-j' \leq e(\tau,P)-1$ we also obtain $\exp_f(\tau,T,j')=\lfloor\frac{e(\tau,T,j')-j'-s}{|H|}\rfloor \leq \lfloor\frac{e(\tau,P)-1-s}{|H|}\rfloor = \exp_f(\tau,P)$. Combining the two upper bound on $\exp_f(\tau,T,j')$, we thus have $\exp_f(\tau,T,j') \leq \min(\exp_f(\tau,P),\lfloor\frac{2\ell-s}{|H|}\rfloor) = k_2$. We next show that $\exp_f(\tau,T,j') \geq k_1$. Consider two cases:

- First, assume $e(\tau,P)-1<\ell$. This implies $\exp_f(\tau,P)=\lfloor\frac{e(\tau,P)-1-s}{|H|}\rfloor\leq\lfloor\frac{\ell-s}{|H|}\rfloor$ and hence $k_1=\min(\exp_f(\tau,P),\lfloor\frac{\ell-s}{|H|}\rfloor)=\exp_f(\tau,P)$. On the other hand, $j'\in\operatorname{Pos}_\ell^{\operatorname{beg}}(P,T)$ implies $\operatorname{lcp}(P,T[j'\mathinner{.\,.} n])\geq\ell$. Thus, $\operatorname{lcp}(P,T[j'\mathinner{.\,.} n])>e(\tau,P)-1$. By Lemma 8.50(1), we therefore obtain $e(\tau,T,j')-j'=e(\tau,P)-1$. Consequently, $\exp_f(\tau,T,j')=\lfloor\frac{e(\tau,T,j')-j'-s}{|H|}\rfloor=\lfloor\frac{e(\tau,P)-1-s}{|H|}\rfloor=\exp_f(\tau,P)=k_1$.
 Let us now assume $e(\tau,P)-1\geq\ell$. Note that this implies $e(\tau,T,j')-j'\geq\ell$, since otherwise
- Let us now assume $e(\tau, P) 1 \ge \ell$. Note that this implies $e(\tau, T, j') j' \ge \ell$, since otherwise by Lemma 8.51(1), we would have $lcp(P, T[j' ... n]) = min(e(\tau, P) 1, e(\tau, T, j') j') < \ell$, which contradicts $j' \in Pos_{\ell}^{beg}(P, T)$. Consequently, $exp_f(\tau, T, j') = \lfloor \frac{e(\tau, T, j') j' s}{|H|} \rfloor \ge \lfloor \frac{\ell s}{|H|} \rfloor$, and hence $k_1 = min(exp_f(\tau, T, j'), \lfloor \frac{\ell s}{|H|} \rfloor) = \lfloor \frac{\ell s}{|H|} \rfloor \le exp_f(\tau, T, j')$.

We have thus proved that $\exp_f(\tau,T,j') \in [k_1 \dots k_2]$. Consider now two cases. If $\exp_f(\tau,T,j') > k_1$, then we have $\exp_f(\tau,T,j') \in (k_1 \dots k_2]$, and hence $j' \in \operatorname{Pos}_{f,\ell}^{\operatorname{mid}-}(P,T)$ by definition. Otherwise (i.e., $\exp_f(\tau,T,j') = k_1$), combining with $\operatorname{lcp}(P,T[j' \dots n]) \geq \ell$ (following from $j' \in \operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T)$) we have $j' \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T)$. Thus, $j' \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T) \cup \operatorname{Pos}_{f,\ell}^{\operatorname{mid}-}(P,T)$, and hence $\operatorname{Pos}_{\ell}^{\operatorname{beg}}(P,T) \subseteq \operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T) \cup \operatorname{Pos}_{f,\ell}^{\operatorname{mid}-}(P,T)$. It therefore remains to show $\operatorname{Pos}_{f,\ell}^{\operatorname{high}f,\ell}(P,T) \subseteq \operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T)$. Let $j' \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}f,\ell}(P,T)$. Consider two

cases. If $k_1 < k_2$, then we immediately obtain $j' \in \operatorname{Pos}^{\operatorname{mid}-}_{f,\ell}(P,T)$, since $\exp_f(\tau,T,j') \in (k_1 \dots k_2]$. Otherwise, we have $\exp_f(\tau,T,j') = k_2 = k_1$, and either $T[j' \dots n] \succeq P$ or $\operatorname{lcp}(P,T[j' \dots n]) \geq 2\ell \geq \ell$. This implies $j' \in \operatorname{Pos}^{\operatorname{low}-}_{f,\ell}(P,T)$. Thus, $\operatorname{Pos}^{\operatorname{high}-}_{f,\ell}(P,T) \subseteq \operatorname{Pos}^{\operatorname{low}-}_{f,\ell}(P,T) \cup \operatorname{Pos}^{\operatorname{mid}-}_{f,\ell}(P,T)$. We have therefore proved that $\operatorname{Pos}^{\operatorname{beg}}_{\ell}(P,T) \cup \operatorname{Pos}^{\operatorname{high}-}_{f,\ell}(P,T) \subseteq \operatorname{Pos}^{\operatorname{low}-}_{f,\ell}(P,T) \cup \operatorname{Pos}^{\operatorname{mid}-}_{f,\ell}(P,T)$.

Definition 8.75. Let f be a necklace-consistent function. Let $\ell \in [16 ... n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. For every $j \in \mathbb{R}(\tau, T)$, we define:

- $\begin{array}{l} \bullet \ \operatorname{Pos}^{\operatorname{low}-}_{f,\ell}(j,T) := \operatorname{Pos}^{\operatorname{low}-}_{f,\ell}(T[j\mathinner{\ldotp\ldotp} n],T), \\ \bullet \ \operatorname{Pos}^{\operatorname{mid}-}_{f,\ell}(j,T) := \operatorname{Pos}^{\operatorname{mid}-}_{f,\ell}(T[j\mathinner{\ldotp\ldotp} n],T), \\ \bullet \ \operatorname{Pos}^{\operatorname{high}-}_{f,\ell}(j,T) := \operatorname{Pos}^{\operatorname{high}-}_{f,\ell}(T[j\mathinner{\ldotp\ldotp} n],T). \end{array}$

The sets $\operatorname{Pos}_{f,\ell}^{\operatorname{low}+}(j,T)$, $\operatorname{Pos}_{f,\ell}^{\operatorname{mid}+}(j,T)$, and $\operatorname{Pos}_{f,\ell}^{\operatorname{high}+}(j,T)$ are defined analogously. We also let

- $\begin{array}{l} \bullet \ \operatorname{Pos}^{\operatorname{low}}_{f,\ell}(j,T) := \operatorname{Pos}^{\operatorname{low}}_{f,\ell}(T[j\mathinner{\ldotp\ldotp} n],T), \\ \bullet \ \operatorname{Pos}^{\operatorname{mid}}_{f,\ell}(j,T) := \operatorname{Pos}^{\operatorname{mid}}_{f,\ell}(T[j\mathinner{\ldotp\ldotp} n],T), \\ \bullet \ \operatorname{Pos}^{\operatorname{high}}_{f,\ell}(j,T) := \operatorname{Pos}^{\operatorname{high}}_{f,\ell}(T[j\mathinner{\ldotp\ldotp} n],T). \end{array}$

Finally, we denote

- $\begin{array}{l} \bullet \ \delta^{\mathrm{low}-}_{f,\ell}(j,T) := |\mathrm{Pos}^{\mathrm{low}-}_{f,\ell}(j,T)|, \\ \bullet \ \delta^{\mathrm{mid}-}_{f,\ell}(j,T) := |\mathrm{Pos}^{\mathrm{mid}-}_{f,\ell}(j,T)|, \\ \bullet \ \delta^{\mathrm{high}-}_{f,\ell}(j,T) := |\mathrm{Pos}^{\mathrm{high}-}_{f,\ell}(j,T)|. \end{array}$

Remark 8.76. Note that all above sets are well-defined since $j \in R(\tau, T)$ implies that T[j ... n] is τ -periodic.

Lemma 8.77. Let f be a necklace-consistent function. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. For every $j \in R(\tau, T)$, it holds:

```
\operatorname{Pos}^{\operatorname{low}-}_{f,\ell}(j,T) = \{j' \in \mathsf{R}^{-}_{f,s,H}(\tau,T) : \exp_{f}(\tau,T,j') = k_1 \ and \ (T[j'\mathinner{\ldotp\ldotp} n] \succeq T[j\mathinner{\ldotp\ldotp} n] \ or \ \operatorname{LCE}_{T}(j,j') \geq \ell)\},
\operatorname{Pos}_{f,\ell}^{\operatorname{mid}^{-}}(j,T) = \{j' \in \mathsf{R}_{f,s,H}^{-}(\tau,T) : \exp_{f}(\tau,T,j') \in (k_{1} \dots k_{2}]\}, \\ \operatorname{Pos}_{f,\ell}^{\operatorname{high}^{-}}(j,T) = \{j' \in \mathsf{R}_{f,s,H}^{-}(\tau,T) : \exp_{f}(\tau,T,j') = k_{2} \text{ and } (T[j' \dots n] \succeq T[j \dots n] \text{ or } \operatorname{LCE}_{T}(j,j') \geq 2\ell)\},
```

where $s = \text{head}_f(\tau, T, j)$, $H = \text{root}_f(\tau, T, j)$, $k_1 = \exp_f^{\text{cut}}(\tau, T, j, \ell)$, and $k_2 = \exp_f^{\text{cut}}(\tau, T, j, 2\ell)$.

Proof. Denote P = T[j ... n]. The claim follows by putting together Definitions 8.70 and 8.75, and observing that (by definition), it holds $\operatorname{root}_f(\tau, P) = \operatorname{root}_f(\tau, T, j) = H$, $\operatorname{head}_f(\tau, P) = \operatorname{head}_f(\tau, T, j) = s$, $\exp_f^{\mathrm{cut}}(\tau, P, \ell) = \exp_f^{\mathrm{cut}}(\tau, T, j, \ell) = k_1, \ \exp_f^{\mathrm{cut}}(\tau, P, 2\ell) = \exp_f^{\mathrm{cut}}(\tau, T, j, 2\ell) = k_2, \ \text{and} \ \operatorname{lcp}(P, T[j' \dots n]) \ge \ell$ (resp. $\operatorname{lcp}(P, T[j' \dots n]) \ge 2\ell$) holds if and only if $\operatorname{LCE}_T(j, j') \ge \ell$ (resp. $\operatorname{LCE}_T(j, j') \ge 2\ell$).

Lemma 8.78. Let f be a necklace-consistent function. Let $\ell \in [16..n)$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. For every $j \in \mathsf{R}^-(\tau,T)$, it holds:

$$\delta_{\ell}^{\mathrm{beg}}(j,T) = \delta_{f,\ell}^{\mathrm{low}-}(j,T) + \delta_{f,\ell}^{\mathrm{mid}-}(j,T) - \delta_{f,\ell}^{\mathrm{high}-}(j,T).$$

Proof. Note that if we let P = T[j ...n], then $\operatorname{type}(\tau,P) = -1$. Thus, by Lemma 8.74, we obtain $\delta_{\ell}^{\operatorname{beg}}(j,T) = \delta_{\ell}^{\operatorname{low}^-}(P,T) = \delta_{f,\ell}^{\operatorname{low}^-}(P,T) + \delta_{f,\ell}^{\operatorname{mid}^-}(P,T) - \delta_{f,\ell}^{\operatorname{high}^-}(P,T) = \delta_{f,\ell}^{\operatorname{low}^-}(j,T) + \delta_{f,\ell}^{\operatorname{mid}^-}(j,T) - \delta_{f,\ell}^{\operatorname{high}^-}(j,T)$.

Lemma 8.79. Let f be a necklace-consistent function. Let $\ell \in [16 \dots n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $P \in \Sigma^+$ be a τ -periodic pattern. For every $i, j \in [1 \dots n]$ such that $T^{\infty}[i \dots i + 7\tau) = T^{\infty}[j \dots j + 7\tau)$, $i \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(\tau, P)$ holds if and only if $j \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(\tau, P)$.

Proof. Denote $H = \operatorname{root}_f(\tau, P)$, $s = \operatorname{head}_f(\tau, P)$, p = |H|, and $k_2 = \exp_f^{\operatorname{cut}}(\tau, P, 2\ell)$. Let $i \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(\tau, P)$. We will prove that $j \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(\tau, P)$ (the proof of the opposite implication follows by symmetry). If i = j, then the claim follows immediately. Let us thus assume $i \neq j$. By Lemma 8.11, we then have $\mathrm{LCE}_T(i,j) \geq 7\tau$. The assumption $i \in \mathrm{Pos}_{f,\ell}^{\mathrm{high}-}(\tau,P)$ implies $i \in \mathsf{R}_{f,s,H}^-(\tau,T)$, $\exp_f(\tau,T,i) = k_2$, and it holds either $T[i\mathinner{.\,.} n] \succeq P$ or $\mathrm{lcp}(P,T[i\mathinner{.\,.} n]) \geq 2\ell$. First, note that by Lemma 8.50(2), we obtain $j \in \mathsf{R}_{f,s,H}(\tau,T)$. Next, recall that $k_2 = \min(\exp_f(\tau,P),\lfloor\frac{2\ell-s}{p}\rfloor) \leq \lfloor\frac{2\ell-s}{p}\rfloor$. Consequently, $e_f^{\mathrm{full}}(\tau,T,i) - i = s + \exp_f(\tau,T,i)p = s + k_2p \leq 2\ell$. This implies $e(\tau,T,i)-i=(e_f^{\mathrm{full}}(\tau,T,i)-i)+(e(\tau,T,i)-e_f^{\mathrm{full}}(\tau,T,i))<(e_f^{\mathrm{full}}(\tau,T,i)-i)+p\leq 2\ell+\lfloor \tau/3\rfloor\leq 7\tau$ (where $2\ell+\lfloor \tau/3\rfloor\leq 7\tau$ follows by $\tau=\lfloor \frac{\ell}{3}\rfloor$ and $\ell\geq 16$). Combining this with $\mathrm{LCE}_T(i,j)\geq 7\tau$ and applying Lemma 8.50(2), we thus have $\exp_f(\tau,T,j)=\exp_f(\tau,T,i)=k_2$ and $\mathrm{type}(\tau,T,j)=\mathrm{type}(\tau,T,i)=-1$, i.e., $j\in \mathbb{R}^-_{f,s,H}(\tau,T)$. It remains to show that it holds $T[j\mathinner{\ldotp\ldotp} n]\succeq P$ or $\mathrm{lcp}(P,T[j\mathinner{\ldotp\ldotp} n])\geq 2\ell$. Recall, that we assumed $T[i\mathinner{\ldotp\ldotp} n]\succeq P$ or $\mathrm{lcp}(P,T[i\mathinner{\ldotp\ldotp} n])\geq 2\ell$. Consider two cases:

- If $lcp(P, T[i ... n]) \ge 2\ell$, then by combining with $lcp(T[i ... n], T[j ... n]) \ge 7\tau$ and $2\ell \le 7\tau$, it follows that $lcp(P, T[j ... n]) \ge \min(lcp(P, T[i ... n]), lcp(T[i ... n], T[j ... n])) \ge \min(2\ell, 7\tau) = 2\ell$.
- Let us now assume that $lcp(P, T[i ... n]) < 2\ell$ and $T[i ... n] \succeq P$. Note that we cannot have P = T[i ... n], since that would imply $n i + 1 = |P| = lcp(P, T[i ... n]) < 2\ell$ which by $2\ell \le 7\tau$ would contradict $n i + 1 \ge 7\tau$ (following from $LCE_T(i, j) \ge 7\tau$). We thus have $T[i ... n] \succ P$, which implies that either P is a proper prefix of T[i ... n], or, letting $\ell' = lcp(P, T[i ... n])$, it holds $\ell' < |P|, i + \ell' \le n$, and $T[i + \ell'] \succ P[1 + \ell']$. Consider two subcases:
 - If P is a proper prefix of T[i ... n], then $|P| = \text{lcp}(P, T[i ... n]) < 2\ell$. Consequently, by $2\ell \le 7\tau$, the assumption $\text{LCE}_T(i, j) \ge 7\tau$ implies that P is also a proper prefix of T[j ... n]. We thus have T[j ... n] > P.
 - Let us now assume that $\ell' < |P|$, $i + \ell' \le n$, and $T[i + \ell'] > P[1 + \ell']$. By $\ell' < 2\ell$, and $LCE_T(i,j) \ge 7\tau$, this implies that $lcp(P,T[j ... n]) = \ell'$ and $T[j + \ell'] = T[i + \ell'] > P[1 + \ell']$. Thus, we again obtain T[j ... n] > P.

8.4.4 The Data Structure

Definitions For every $k \in [4..\lceil \log n \rceil)$, denote $\ell_k = 2^k$, $\tau_k = \lfloor \frac{\ell_k}{3} \rfloor$, $f_k := f_{\tau_k,T}$ (Definition 8.64), $n_{\text{runs},k} = |\mathcal{I}(\text{comp}_{14\tau_k}(\mathsf{R}(\tau_k,T),T))|$ (Definitions 4.2 and 4.10), and let $A_{\text{root},k}[1..n_{\text{runs},k}]$ be such that for $j \in [1..n_{\text{runs},k}]$ it holds $A_{\text{root},k}[j] = (\text{head}_{f_k}(\tau_k,T,p_j),|\text{root}_{f_k}(\tau_k,T,p_j)|)$, where $(p_j,t_j)_{j \in [1..n_{\text{runs},k}]} = \mathcal{I}(\text{comp}_{14\tau_k}(\mathsf{R}(\tau_k,T),T))$.

Components The data structure, denoted CompSAPeriodic(T), consists of two parts. The first part consists of the following four components:

- 1. The index core CompSACore(T) (Section 8.2.2). It needs $\mathcal{O}(\delta(T)\log \frac{n\log \sigma}{\delta(T)\log n})$ space.
- 2. For $k \in [4.. \lceil \log n \rceil)$, we store the array $A_{\text{root},k}[1..n_{\text{runs},k}]$ in plain form using $\mathcal{O}(n_{\text{runs},k})$ space. By the same analysis as in Section 8.2.2, the total space used by the arrays is bounded by

$$\mathcal{O}(\sum_{k \in [4..\lceil \log n \rceil)} n_{\text{runs},k}) = \mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n}).$$

3. For every $k \in [4..\lceil \log n \rceil)$ and $H \in \Sigma^+$ such that $\operatorname{Seed}_{f_k,H}^-(\tau_k,T) \neq \emptyset$ (Definition 8.60), we store a data structure from Proposition 6.6 for $q = 7\tau_k$ and $P = \operatorname{Seed}_{f_k,H}^-(\tau_k,T)$. By Proposition 6.6, the size of a single structure is $\mathcal{O}(|\operatorname{Seed}_{f_k,H}^-(\tau_k,T)|)$. By Corollary 8.62, the total size of all the structures is thus

$$\mathcal{O}(\sum_{k \in [4..\lceil \log n \rceil)} \sum_{H \in \Sigma^+} |\mathrm{Seed}_{f_k, H}^-(\tau_k, T)|) = \mathcal{O}(\sum_{k \in [4..\lceil \log n \rceil)} n_{\mathrm{runs}, k})$$
$$= \mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n}).$$

To access the structures, for every $k \in [4..\lceil \log n \rceil)$, we store an array $A_{\text{ptr},k}[1..n_{\text{runs},k}]$ such that letting $(p_j, t_j)_{j \in [1..n_{\text{runs},k}]} = \mathcal{I}(\text{comp}_{14\tau_k}(\mathsf{R}(\tau_k, T), T))$, $A_{\text{ptr},k}[j]$ stores the pointer to the structure for $P = \text{Seed}^-_{f_k,H}(\tau_k,T)$, where $H = \text{root}_{f_k}(\tau_k,T,p_j)$ (or a null pointer, if $\text{Seed}^-_{f_k,H}(\tau_k,T) = \emptyset$). In total, the arrays need $\mathcal{O}(\sum_{k \in [4..\lceil \log n \rceil}) n_{\text{runs},k}) = \mathcal{O}(\delta(T) \log \frac{n \log \sigma}{\delta(T) \log n})$ space.

4. For $k \in [4..\lceil \log n \rceil]$ and $H \in \Sigma^+$ such that $\text{Seed}^-_{f_k,H}(\tau_k,T) \neq \emptyset$, we store a data structure from Proposition 7.3 for a T and T are small T and T are T are T and T and T are T are T and T are T a

4. For $k \in [4 ... \lceil \log n \rceil)$ and $H \in \Sigma^+$ such that $\operatorname{Seed}_{f_k,H}^-(\tau_k,T) \neq \emptyset$, we store a data structure from Proposition 7.3 for $q = 7\tau_k$ and $P = \operatorname{Seed}_{f_k,H}^-(\tau_k,T)$, with h = |H|. The assumptions from Proposition 7.3 are satisfied since the labels in WeightedIntervals_{7 τ}($\operatorname{Seed}_{f_k,H}^-(\tau_k,T)$, T) are unique by Lemma 8.63. By Proposition 7.3, the size of a single structure is $\mathcal{O}(|\operatorname{Seed}_{f_k,H}^-(\tau_k,T)|)$. In total the structures need $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ (see above). To access the structures, we store an array $A'_{\operatorname{ptr},k}[1 ... n_{\operatorname{runs},k}]$ such that, letting $(p_j,t_j)_{j\in[1...n_{\operatorname{runs},k}]} = \mathcal{I}(\operatorname{comp}_{14\tau_k}(\mathsf{R}(\tau_k,T),T))$, $A'_{\operatorname{ptr},k}[j]$ stores the pointer to the structure

for $P = \operatorname{Seed}_{f_k,H}^-(\tau_k,T)$, where $H = \operatorname{root}_{f_k}(\tau_k,T,p_j)$ (or a null pointer, if $\operatorname{Seed}_{f_k,H}^-(\tau_k,T) = \emptyset$). In total, the arrays need $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ space (see above).

The second part of the structure consists of the same components as above, except in all definitions the standard lexicographic order \preceq is replaced with the inverted lexicographic order \preceq_{inv} (Definition 8.2). No other orders are changed. As a result, e.g., (1) In Definition 8.60, $\mathsf{R}_{f,H}(\tau,T)$ is replaced with $\mathsf{R}_{f,H}^+(\tau,T)$, and (2) In the structures for range queries (Proposition 6.6) we use \preceq_{inv} instead of \preceq as the order on the second coordinate (while the order on the first coordinate remains the same). See Remark 8.71 for the justification. In total, CompSAPeriodic(T) needs $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ space.

Remark 8.80. Note that in the definition of $A_{\text{ptr},k}$ (resp. $A'_{\text{ptr},k}$), we store in $A_{\text{ptr},k}[j]$ (resp. $A'_{\text{ptr},k}[j]$) the pointer to the structure from Proposition 6.6 (resp. Proposition 7.3) for $P = \text{Seed}^-_{f_k,H}(\tau_k,T)$, where $H = \text{root}_{f_k}(\tau_k,T,p_j)$ and $\text{Seed}^-_{f_k,H}(\tau_k,T) \neq \emptyset$, even if $\text{type}(\tau_k,T,p_j) = +1$. This is used, e.g., in Proposition 8.101.

8.4.5 Basic Combinatorial Properties

Weighted Range Queries

Lemma 8.81. Let $\tau \in [1..\lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Let $H \in \Sigma^+$ and $\mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f,H}^-(\tau,T),T)$ (Definitions 6.5 and 8.60). For any $x_l \in [0..7\tau]$ and $y_l, y_u \in \Sigma^*$, it holds:

- $\begin{aligned} 1. \ \text{weight-count}_{\mathcal{P}}(x_l, n, y_l, y_u) &= \\ |\{j \in \mathsf{R}_{f,H}'^{-}(\tau, T) : x_l \leq e_f^{\mathrm{full}}(\tau, T, j) j, \, and \, y_l \preceq T^{\infty}[e_f^{\mathrm{full}}(\tau, T, j) \ldots e_f^{\mathrm{full}}(\tau, T, j) + 7\tau) \prec y_u\}|, \end{aligned}$
- $\begin{aligned} & \text{2. weight-count} \underset{\mathcal{T}}{\preceq}(x_l,n,y_u) = \\ & |\{j \in \mathsf{R}_{f,H}'^{-}(\tau,T): x_l \leq e_f^{\mathrm{full}}(\tau,T,j) j, and \ T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j) \ldots e_f^{\mathrm{full}}(\tau,T,j) + 7\tau) \preceq y_u\}|, \end{aligned}$
- $\begin{array}{l} \textit{3.} \ \ \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l,n) = \\ |\{j \in \mathsf{R}'^-_{f,H}(\tau,T) : x_l \leq e^{\mathrm{full}}_f(\tau,T,j) j\}|. \end{array}$

Proof. 1. Denote $\mathsf{R}_{\mathrm{comp}} = \mathrm{comp}_{14\tau}(\mathsf{R}_{f,H}^{-}(\tau,T),T)$, and recall that we have $\mathrm{Seed}_{f,H}^{-}(\tau,T) = \{(e_f^{\mathrm{full}}(\tau,T,j), \min(e_f^{\mathrm{full}}(\tau,T,j)-b(\tau,T,j),7\tau)): j \in \mathsf{R}_{\mathrm{comp}}\}$ (Definition 8.60). The proof consists of four steps labeled (a) through (d).

(a) Denote

$$Q = \{ T^{\infty}[e_f^{\text{full}}(\tau, T, i) - 7\tau ... e_f^{\text{full}}(\tau, T, i) + 7\tau) : i \in \mathsf{R}_{\text{comp}}, x_l \le e_f^{\text{full}}(\tau, T, i) - b(\tau, T, i), \text{ and } y_l \le T^{\infty}[e_f^{\text{full}}(\tau, T, i) ... e_f^{\text{full}}(\tau, T, i) + 7\tau) \prec y_u \},$$

$$A = \{ j \in [1 ... n] : T^{\infty}[j - 7\tau ... j + 7\tau) \in Q \}.$$

In the first step, we prove that weight-count $p(x_l, n, y_l, y_u) = |A|$. Let $\mathcal{R} = \{(x, y, w, \ell) \in \mathcal{P} : x_l \leq x < n \text{ and } y_l \leq y \prec y_u\}$. First, observe that for every $p = (x_p, y_p, w_p, \ell_p) \in \mathcal{R}$, it holds $T^{\infty}[\ell_p - 7\tau \dots \ell_p + 7\tau) \in Q$. To see this, note that $p \in \text{IntStrPoints}_{7\tau}(\text{Seed}_{f,H}^{-}(\tau,T),T)$. By definition of $\text{Seed}_{f,H}^{-}(\tau,T)$, there exists $i_p \in \mathcal{R}_{\text{comp}}$ such that $x_p = \min(e_f^{\text{full}}(\tau,T,i_p) - b(\tau,T,i_p),7\tau)$, $y_p = T^{\infty}[e_f^{\text{full}}(\tau,T,i_p) \dots e_f^{\text{full}}(\tau,T,i_p) + 7\tau)$, and ℓ_p satisfies $T^{\infty}[\ell_p - 7\tau \dots \ell_p + 7\tau) = T^{\infty}[e_f^{\text{full}}(\tau,T,i_p) - 7\tau \dots e_f^{\text{full}}(\tau,T,i_p) + 7\tau)$. Thus, $p \in \mathcal{R}$ implies $x_l \leq x_p = \min(e_f^{\text{full}}(\tau,T,i_p) - b(\tau,T,i_p),7\tau) \leq e_f^{\text{full}}(\tau,T,i_p) - b(\tau,T,i_p)$ and $y_l \leq y_p = T^{\infty}[e_f^{\text{full}}(\tau,T,i_p) \dots e_f^{\text{full}}(\tau,T,i_p) + 7\tau) \prec y_u$. By definition we thus have $T^{\infty}[e_f^{\text{full}}(\tau,T,i_p) - 7\tau \dots e_f^{\text{full}}(\tau,T,i_p) + 7\tau) \in Q$. Since $T^{\infty}[e_f^{\text{full}}(\tau,T,i_p) - 7\tau \dots e_f^{\text{full}}(\tau,T,i_p) + 7\tau) \in Q$. Let $g: \mathcal{R} \to Q$ be defined so that for every $p = (x,y,w,\ell) \in \mathcal{R}$, we let $g(p) = T^{\infty}[\ell - 7\tau \dots \ell + 7\tau)$. We prove that g is a bijection.

• Let $p_1, p_2 \in \mathcal{R}$ be such that $g(p_1) = g(p_2)$. We show that $p_1 = p_2$. For $k \in \{1, 2\}$, denote $p_k = (x_k, y_k, w_k, \ell_k)$ and let $i_k \in \mathsf{R}_{\text{comp}}$ be such that $x_k = \min(e_f^{\text{full}}(\tau, T, i_k) - b(\tau, T, i_k), 7\tau)$, $y_k = T^{\infty}[e_f^{\text{full}}(\tau, T, i_k) \dots e_f^{\text{full}}(\tau, T, i_k) + 7\tau)$, and $T^{\infty}[\ell_k - 7\tau \dots \ell_k + 7\tau) = T^{\infty}[e_f^{\text{full}}(\tau, T, i_k) - 7\tau \dots e_f^{\text{full}}(\tau, T, i_k) + 7\tau)$. Let also $S_k := T^{\infty}[\ell_k - 7\tau \dots \ell_k + 7\tau) = T^{\infty}[e_f^{\text{full}}(\tau, T, i_k) - 7\tau \dots e_f^{\text{full}}(\tau, T, i_k) + 7\tau)$. By the assumption, $S_1 = g(p_1) = g(p_2) = S_2$. Consequently, it holds $\ell_1 = \min\{j \in [1 \dots n] : T^{\infty}[j - 7\tau \dots j + 1]\}$

- $\begin{array}{l} 7\tau) = S_1 \} = \min \{ j \in [1 \ldots n] : T^{\infty}[j 7\tau \ldots j + 7\tau) = S_2 \} = \ell_2. \text{ Similarly, } w_1 = | \{ j \in [1 \ldots n] : T^{\infty}[j 7\tau \ldots j + 7\tau) = S_1 \} | = | \{ j \in [1 \ldots n] : T^{\infty}[j 7\tau \ldots j + 7\tau) = S_2 \} | = w_2. \text{ We also have } y_1 = T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i_1) \ldots e_f^{\mathrm{full}}(\tau, T, i_1) + 7\tau) = S_1(7\tau \ldots 14\tau] = S_2(7\tau \ldots 14\tau] = T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i_2) \ldots e_f^{\mathrm{full}}(\tau, T, i_2) + 7\tau) = y_2. \text{ Lastly, observe that due to } T[n] = T^{\infty}[0] \text{ being unique in } T, \text{ for every } i \in \mathsf{R}_{f,H}(\tau, T), \text{ letting } Z = T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) 7\tau \ldots e_f^{\mathrm{full}}(\tau, T, i)), \text{ we have } \min(e_f^{\mathrm{full}}(\tau, T, i) b(\tau, T, i), 7\tau) = |H| + |\cos(Z, Z[1 \ldots 7\tau |H|]). \text{ Thus, } x_1 = \min(e_f^{\mathrm{full}}(\tau, T, i_1) b(\tau, T, i_1), 7\tau) = |H| + |\cos(S_1[1 \ldots 7\tau], S_1[1 \ldots 7\tau |H|]) = |H| + |\cos(S_2[1 \ldots 7\tau], S_2[1 \ldots 7\tau |H|]) = \min(e_f^{\mathrm{full}}(\tau, T, i_2) b(\tau, T, i_2), 7\tau) = x_2. \text{ We have thus proved } p_1 = p_2. \end{array}$
- Let us now consider any $X \in Q$. Then, there exists $i \in \mathsf{R}_{\mathsf{comp}}$ such that $x_l \leq e_f^{\mathsf{full}}(\tau, T, i) b(\tau, T, i)$, $y_l \leq T^{\infty}[e_f^{\mathsf{full}}(\tau, T, i) \dots e_f^{\mathsf{full}}(\tau, T, i) + 7\tau) \prec y_u$, and $T^{\infty}[e_f^{\mathsf{full}}(\tau, T, i) 7\tau \dots e_f^{\mathsf{full}}(\tau, T, i) = 7\tau) = X$. Note that by definition of $\mathsf{Seed}_{f,H}^-(\tau, T)$, $i \in \mathsf{R}_{\mathsf{comp}}$ implies that $(e_f^{\mathsf{full}}(\tau, T, i), \mathsf{min}(e_f^{\mathsf{full}}(\tau, T, i) b(\tau, T, i), 7\tau)) \in \mathsf{Seed}_{f,H}^-(\tau, T)$. Consequently, there exists $p = (x, y, w, \ell) \in \mathcal{P}$ such that $x = \min(e_f^{\mathsf{full}}(\tau, T, i) b(\tau, T, i), 7\tau)$, $y = T^{\infty}[e_f^{\mathsf{full}}(\tau, T, i) \dots e_f^{\mathsf{full}}(\tau, T, i) + 7\tau)$, and ℓ satisfies $T^{\infty}[\ell 7\tau \dots \ell + 7\tau) = T^{\infty}[e_f^{\mathsf{full}}(\tau, T, i) 7\tau \dots e_f^{\mathsf{full}}(\tau, T, i) + 7\tau) = X$. By $x_l \leq 7\tau$, the inequality $x_l \leq e_f^{\mathsf{full}}(\tau, T, i) b(\tau, T, i)$ implies $x_l \leq \min(e_f^{\mathsf{full}}(\tau, T, i) b(\tau, T, i), 7\tau) = x$. On the other hand, $x \leq e_f^{\mathsf{full}}(\tau, T, i) b(\tau, T, i) \leq n 1 < n$ holds by the uniqueness of T[n] in T. Finally,note that we then also have $y_l \leq y = T^{\infty}[e_f^{\mathsf{full}}(\tau, T, i) \dots e_f^{\mathsf{full}}(\tau, T, i) + 7\tau) \prec y_u$. Thus, it holds $p \in \mathcal{R}$. Since, as noted above, ℓ satisfies $T^{\infty}[\ell 7\tau \dots \ell + 7\tau) = X$, we thus obtain $g(p) = T^{\infty}[\ell 7\tau \dots \ell + 7\tau) = X$.

We have thus proved that g is a bijection. Thus, for every $(x_1,y_1,w_1,\ell_1), (x_2,y_2,w_2,\ell_2) \in \mathcal{R}, (x_1,y_1,w_1,\ell_1) \neq (x_2,y_2,w_2,\ell_2)$ implies $T^{\infty}[\ell_1-7\tau \mathinner{...}\ell_1+7\tau) \neq T^{\infty}[\ell_2-7\tau \mathinner{...}\ell_2+7\tau)$. Observe now that in Definition 6.5, we can equivalently set $c(i)=|\{i'\in[1\mathinner{...}n]:T^{\infty}[i'-q\mathinner{...}i'+q)=T^{\infty}[m(i)-q\mathinner{...}m(i)+q)\}|$. On the other hand, by the above, letting $P_k=\{j\in[1\mathinner{...}n]:T^{\infty}[j-7\tau\mathinner{...}j+7\tau)=T^{\infty}[\ell_k-7\tau\mathinner{...}\ell_k+7\tau)\}$ for $k\in\{1,2\}$, we have $P_1\cap P_2=\emptyset$. Combining that, we obtain weight-count $_{\mathcal{P}}(x_l,n,y_l,y_u)=\sum_{(x,y,w,\ell)\in\mathcal{R}}w=\sum_{(x,y,w,\ell)\in\mathcal{R}}|\{j\in[1\mathinner{...}n]:T^{\infty}[j-7\tau\mathinner{...}j+7\tau)=T^{\infty}[\ell-7\tau\mathinner{...}\ell+7\tau)\}|=|\{j\in[1\mathinner{...}n]:T^{\infty}[j-7\tau\mathinner{...}j+7\tau)\in Q\}|=|A|.$ (b) Denote

$$Q' = \{ T^{\infty}[e_f^{\text{full}}(\tau, T, i) - 7\tau \dots e_f^{\text{full}}(\tau, T, i) + 7\tau) : i \in \mathsf{R}_{f, H}^{-}(\tau, T), x_l \le e_f^{\text{full}}(\tau, T, i) - i, \text{ and } y_l \le T^{\infty}[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + 7\tau) \prec y_u \}.$$

In the second step, we show that Q = Q'.

- First, consider any $X \in Q'$. Then, there exists $i \in \mathsf{R}^-_{f,H}(\tau,T)$ such that $X = T^\infty[e_f^{\mathrm{full}}(\tau,T,i) 7\tau \dots e_f^{\mathrm{full}}(\tau,T,i) + 7\tau)$, $x_l \leq e_f^{\mathrm{full}}(\tau,T,i) i$, and $y_l \leq T^\infty[e_f^{\mathrm{full}}(\tau,T,i) \dots e_f^{\mathrm{full}}(\tau,T,i) + 7\tau) \prec y_u$. Let $i' = \max(i,e_f^{\mathrm{full}}(\tau,T,i) 7\tau)$. Note that then $[i\dots i'] \subseteq \mathsf{R}(\tau,T)$ and thus by Lemma 8.48, $i' \in \mathsf{R}^-_{f,H}(\tau,T)$ and $e_f^{\mathrm{full}}(\tau,T,i') = e_f^{\mathrm{full}}(\tau,T,i)$. By $x_l \leq 7\tau$, we then also have $e_f^{\mathrm{full}}(\tau,T,i') i' \in [x_l \dots 7\tau]$. Let $j = \min\{t \in [1\dots n] : T^\infty[t 7\tau \dots t + 7\tau) = X\}$ and let j' be such that $e_f^{\mathrm{full}}(\tau,T,i') i' = j j'$. We prove the following properties of j':
 - First, we show that $j' \in \mathsf{R}^-_{f,H}(\tau,T)$ and $e^{\mathrm{full}}_f(\tau,T,j') = j$. Observe, that letting $\delta = j j'$ and $\ell' = (j-j')+\tau$, we have $\delta \in [0\ldots e(\tau,T,i')-i']$, $e(\tau,T,i')-i' = (e(\tau,T,i')-e^{\mathrm{full}}_f(\tau,T,i'))+(e^{\mathrm{full}}_f(\tau,T,i')-i')+(e^{\mathrm{full}}_f(\tau,$
 - Second, we prove that $y_l \leq T^{\infty}[e_f^{\text{full}}(\tau, T, j') \dots e_f^{\text{full}}(\tau, T, j') + 7\tau) \prec y_u$. For this, it suffices to note that $T^{\infty}[e_f^{\text{full}}(\tau, T, j') \dots e_f^{\text{full}}(\tau, T, j') + 7\tau) = T^{\infty}[j \dots j + 7\tau) = X(7\tau \dots 14\tau]$. On the other hand, $X(7\tau \dots 14\tau] = T^{\infty}[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + 7\tau)$. Thus, by the assumption about the string $T^{\infty}[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + 7\tau)$, we obtain $y_l \leq T^{\infty}[e_f^{\text{full}}(\tau, T, j') \dots e_f^{\text{full}}(\tau, T, j') + 7\tau) \prec y_u$.

 Finally, we prove $j' \in \mathsf{C}(14\tau, T)$ (Construction 4.6). For this, by Lemma 4.5 it suffices to prove
 - Finally, we prove $j' \in C(14\tau, T)$ (Construction 4.6). For this, by Lemma 4.5 it suffices to prove that there exists $p \in [1 ... n]$ such that $p = \min \operatorname{Occ}_{14\tau}(p, T)$ and $j' \in [p ... p + 14\tau)$. If $j \leq 14\tau$, then by $j' \leq j$ it suffices to take p = 1. Let us thus assume $j > 14\tau$ and let $p = j 7\tau$. To show that $p = \min \operatorname{Occ}_{14\tau}(p, T)$, note that by definition of j, for every $t \in [1 ... j)$, we have $T^{\infty}[t 7\tau ... t + 7\tau) \neq X$. Thus, for every $p' \in [1 ... p)$, it holds $T^{\infty}[p' ... p' + 14\tau) \neq X$. By

 $T^{\infty}[p ... p + 14\tau) = X$, we therefore obtain $p = \min \operatorname{Occ}_{14\tau}(p, T)$. It remains to note that j' < j and $j' = j - (e_f^{\text{full}}(\tau, T, i') - i') \ge j - 7\tau$. Thus, $j' \in [j - 7\tau ... j) \subseteq [p ... p + 14\tau)$.

By combining the above properties, we obtain that $j' \in \mathsf{R}^-_{f,H}(\tau,T) \cap \mathsf{C}(14\tau,T) = \mathsf{R}_{\mathrm{comp}}, \ e_f^{\mathrm{full}}(\tau,T,j') - b(\tau,T,j') \geq e_f^{\mathrm{full}}(\tau,T,j') - j' = j - j' = e_f^{\mathrm{full}}(\tau,T,i') - i' \geq x_l, \ \text{and} \ y_l \preceq T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j') \dots e_f^{\mathrm{full}}(\tau,T,j') + 7\tau) \prec y_u.$ Thus, by definition of Q, we have $T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j') - 7\tau \dots e_f^{\mathrm{full}}(\tau,T,j') + 7\tau) \in Q$. Recalling, that $T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j') - 7\tau \dots e_f^{\mathrm{full}}(\tau,T,j') + 7\tau) = T^{\infty}[j - 7\tau \dots j + 7\tau] = X$, we thus obtain $X \in Q$.

- Let us now consider $X \in Q$. By definition, then there exists $i \in R_{\text{comp}}$ such that $x_l \leq e_f^{\text{tull}}(\tau, T, i) b(\tau, T, i), \ y_l \leq T^{\infty}[e_f^{\text{tull}}(\tau, T, i) \dots e_f^{\text{tull}}(\tau, T, i) + 7\tau) \leq y_u, \ \text{and} \ T^{\infty}[e_f^{\text{tull}}(\tau, T, i) 7\tau \dots e_f^{\text{tull}}(\tau, T, i) + 7\tau) = X$. By $R_{\text{comp}} = C(14\tau, T) \cap R_{f,H}^-(\tau, T) \subseteq R_{f,H}^-(\tau, T), \ \text{we have} \ i \in R_{f,H}^-(\tau, T).$ Denote $i' = b(\tau, T, i)$. By Lemma 8.48, we have $i' \in R_{f,H}^-(\tau, T)$ and $e_f^{\text{tull}}(\tau, T, i') = e_f^{\text{tull}}(\tau, T, i)$. Thus, it holds $T^{\infty}[e_f^{\text{full}}(\tau, T, i') \dots e_f^{\text{full}}(\tau, T, i') + 7\tau) = T^{\infty}[e_f^{\text{tull}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + 7\tau)$ and hence $y_l \leq T^{\infty}[e_f^{\text{tull}}(\tau, T, i') \dots e_f^{\text{tull}}(\tau, T, i') + 7\tau) \leq y_u$. By $e_f^{\text{tull}}(\tau, T, i) \geq x_l$, we also obtain $e_f^{\text{tull}}(\tau, T, i') i' = e_f^{\text{tull}}(\tau, T, i) b(\tau, T, i) \geq x_l$. We have thus proved that $X = T^{\infty}[e_f^{\text{tull}}(\tau, T, i) 7\tau \dots e_f^{\text{tull}}(\tau, T, i) + 7\tau) = T^{\infty}[e_f^{\text{tull}}(\tau, T, i') 7\tau \dots e_f^{\text{tull}}(\tau, T, i') + 7\tau) \in Q'$.
- (c) Denote

 $A' = \{e_f^{\mathrm{full}}(\tau, T, i) : i \in \mathsf{R}_{f, H}'^-(\tau, T), x_l \leq e_f^{\mathrm{full}}(\tau, T, i) - i, \text{ and } y_l \preceq T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) \ldots e_f^{\mathrm{full}}(\tau, T, i) + 7\tau) \prec y_u\}.$

In the third step, we prove that A = A'.

- Consider any $j \in A'$. Then, there exists $j' \in \mathsf{R}'_{f,H}(\tau,T)$ such that $j = e_f^{\mathrm{full}}(\tau,T,j'), x_l \leq e_f^{\mathrm{full}}(\tau,T,j') j',$ and $y_l \leq T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j') \ldots e_f^{\mathrm{full}}(\tau,T,j') + 7\tau) \prec y_u$. To show that $j \in A$, we need to prove that $T^{\infty}[j-7\tau \ldots j+7\tau) \in Q'$ (we use here that Q = Q'). This in turn requires showing that there exists $t \in \mathsf{R}^-_{f,H}(\tau,T)$ such that $x_l \leq e_f^{\mathrm{full}}(\tau,T,t) t, y_l \leq T^{\infty}[e_f^{\mathrm{full}}(\tau,T,t) \ldots e_f^{\mathrm{full}}(\tau,T,t) + 7\tau) \prec y_u$, and $T^{\infty}[e_f^{\mathrm{full}}(\tau,T,t) 7\tau \ldots e_f^{\mathrm{full}}(\tau,T,t) + 7\tau) = T^{\infty}[j-7\tau \ldots j+7\tau)$. Let t=j'. Then, we indeed have $t \in \mathsf{R}'_{f,H}(\tau,T) \subseteq \mathsf{R}^-_{f,H}(\tau,T), e_f^{\mathrm{full}}(\tau,T,t) t = e_f^{\mathrm{full}}(\tau,T,j') j' \geq x_l$ and $T^{\infty}[e_f^{\mathrm{full}}(\tau,T,t) \ldots e_f^{\mathrm{full}}(\tau,T,t) + 7\tau) = T^{\infty}[e_f^{\mathrm{full}}(\tau,T,t) \ldots e_f^{\mathrm{full}}(\tau,T,t) + 7\tau)$ (which implies $y_l \leq T^{\infty}[e_f^{\mathrm{full}}(\tau,T,t) \ldots e_f^{\mathrm{full}}(\tau,T,t) + 7\tau) \prec y_u$). Lastly, $T^{\infty}[e_f^{\mathrm{full}}(\tau,T,t) 7\tau \ldots e_f^{\mathrm{full}}(\tau,T,t) + 7\tau) = T^{\infty}[j-7\tau \ldots j+7\tau)$. Thus, $j \in A$.
- Let us now take $j \in A$. Then, $j \in [1 \dots n]$ and $T^{\infty}[j-7\tau \dots j+7\tau) \in Q'$ (we again use that Q = Q'), which in turn implies that there exists a position $i \in \mathsf{R}_{f,H}^-(\tau,T)$ such that $T^{\infty}[e_f^{\mathrm{full}}(\tau,T,i)-7\tau \dots e_f^{\mathrm{full}}(\tau,T,i)+7\tau) = T^{\infty}[j-7\tau \dots j+7\tau)$, $x_l \leq e_f^{\mathrm{full}}(\tau,T,i)-i$, and $y_l \preceq T^{\infty}[e_f^{\mathrm{full}}(\tau,T,i)\dots e_f^{\mathrm{full}}(\tau,T,i)+7\tau) \prec y_u$. Let $i' = \max(i,e_f^{\mathrm{full}}(\tau,T,i)-7\tau)$. We proceed analogously as in the second step. First, note that then $[i \dots i'] \subseteq \mathsf{R}(\tau,T)$ and thus by Lemma 8.48, $i' \in \mathsf{R}_{f,H}^-(\tau,T)$ and $e_f^{\mathrm{full}}(\tau,T,i') = e_f^{\mathrm{full}}(\tau,T,i)$. By $x_l \leq 7\tau$, we then also have $e_f^{\mathrm{full}}(\tau,T,i')-i' \in [x_l \dots 7\tau]$. Let j' be such that $e_f^{\mathrm{full}}(\tau,T,i')-i' = j-j'$. Letting $\delta = j-j'$ and $\ell' = (j-j')+\tau$, we have $\delta \in [0\dots e(\tau,T,i')-i']$ and $e(\tau,T,i')-i' < \ell'$. Moreover, $T^{\infty}[e_f^{\mathrm{full}}(\tau,T,i')-7\tau \dots e_f^{\mathrm{full}}(\tau,T,i')+7\tau) = T^{\infty}[j-7\tau \dots j+7\tau)$ implies $T^{\infty}[j-\delta \dots j-\delta+\ell') = T^{\infty}[j'\dots j+\tau) = T^{\infty}[i'\dots e_f^{\mathrm{full}}(\tau,T,i')+\tau) = T^{\infty}[i'\dots i'+\ell')$. By Lemma 8.55, we thus obtain $j' \in \mathsf{R}_{f,H}^-(\tau,T)$, and $e_f^{\mathrm{full}}(\tau,T,j') = j' + (e_f^{\mathrm{full}}(\tau,T,i')-i') = j$. Let us now define $j'' = b(\tau,T,j')$. By definition, $j'' \in \mathsf{R}'(\tau,T)$. On the other hand, by Lemma 8.48, $j'' \in \mathsf{R}_{f,H}^-(\tau,T)$. Thus, $j'' \in \mathsf{R}'_{f,H}^-(\tau,T)$. By Lemma 8.48, we also have $e_f^{\mathrm{full}}(\tau,T,j'') = e_f^{\mathrm{full}}(\tau,T,j') = j$. Next, note that $j'' = b(\tau,T,j') \leq j'$ and our earlier assumptions imply $e_f^{\mathrm{full}}(\tau,T,j'') j'' = e_f^{\mathrm{full}}(\tau,T,j') j'' \geq e_f^{\mathrm{full}}(\tau,T,j') j' = j j' = e_f^{\mathrm{full}}(\tau,T,i') i' \geq x_l$. Lastly, we also have $T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j'') \dots e_f^{\mathrm{full}}(\tau,T,j'') + \tau \tau = T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j') \dots e_f^{\mathrm{full}}(\tau,T,j'') + \tau \tau = T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j') \dots e_f^{\mathrm{full}}(\tau,T,j'') + \tau \tau = T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j'') \dots e_f^{\mathrm{full}}(\tau,T,j'') + \tau \tau = T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j'') \dots e_f^{\mathrm{full}}(\tau,T,j'') + \tau \tau = T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j'') \dots e_f^{\mathrm{full}}(\tau,T,j'') + \tau \tau \to y_l$. We have thus proved that $j \in A'$
- (d) Denote

$$A'' = \{i \in \mathsf{R}_{f,H}'^{-}(\tau,T) : x_l \leq e_f^{\mathrm{full}}(\tau,T,i) - i \text{ and } y_l \preceq T^{\infty}[e_f^{\mathrm{full}}(\tau,T,i) \ldots e_f^{\mathrm{full}}(\tau,T,i) + 7\tau) \prec y_u\}.$$

In this step, we put everything together. By Lemma 8.54, for every $i, i' \in R'(\tau, T)$, $i \neq i'$ implies $e_f^{\text{full}}(\tau, T, i) \neq e_f^{\text{full}}(\tau, T, i')$. This implies |A'| = |A''|. Combining this with (a) and (c), we thus obtain weight-count_P $(x_l, n, y_l, y_u) = |A| = |A'| = |A''|$, i.e., the claim.

- 2. The proof is analogous, except we replace the condition $y_l \preceq T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) \dots e_f^{\mathrm{full}}(\tau, T, i) + 7\tau) \prec y_u$ in the definition of Q, Q', A', and A'' with the condition $T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) \dots e_f^{\mathrm{full}}(\tau, T, i) + 7\tau) \preceq y_u$.

 3. The proof is again analogous, except we remove the condition $y_l \preceq T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) \dots e_f^{\mathrm{full}}(\tau, T, i) + 7\tau) \prec f_{\mathrm{full}}(\tau, T, i)$.
- y_u in the definition of Q, Q', A', and A''.

Lemma 8.82. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Let $H \in \Sigma^+$, p = |H|, $s \in [0 ... p)$, $k_{\min} = \lceil \frac{3\tau - 1 - s}{p} \rceil - 1$, $k_{\max} = \lfloor \frac{7\tau - s}{p} \rfloor$, and $\mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f,H}^-(\tau,T),T)$. Then:

- $\bullet \ \ For \ every \ k \in [1\mathinner{\ldotp\ldotp} k_{\min}], \ it \ \ holds \ \ \text{weight-count}_{\mathcal{P}}(s+kp,n) \geq |\mathsf{R}^-_{f,s,k,H}(\tau,T)|,$
- For every $k \in (k_{\min} ... k_{\max}]$, it holds weight-count_P $(s + kp, n) = |R_{f.s.k.H}^{-}(\tau, T)|$.

Proof. By combining Lemma 8.59 and Lemma 8.81, it holds for $k \in [1..k_{\min}]$ that:

$$\begin{split} |\mathsf{R}^-_{f,s,k,H}(\tau,T)| &\leq |\{j \in \mathsf{R}'^-_{f,H}(\tau,T) : s + kp \leq e_f^{\mathrm{full}}(\tau,T,j) - j\}| \\ &= \mathsf{weight\text{-}count}_{\mathcal{P}}(s + kp,n). \end{split}$$

The proof for $k \in (k_{\min} ... k_{\max}]$ follows analogously, except the inequality in the first line is replaced with an equality. Note that Lemma 8.81 requires that $s + kp \le 7\tau$, which holds here, since $k \le k_{\text{max}}$ implies $s + kp \le s + \lfloor \frac{7\tau - s}{p} \rfloor p \le 7\tau.$

Lemma 8.83. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Let $H \in \Sigma^+$, p = |H|, $s \in [0 ... p)$, $k_{\min} = \lceil \frac{3\tau - 1 - s}{p} \rceil - 1, \ k_{\max} = \lceil \frac{7\tau - s}{p} \rceil, \ and \ \mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f,H}^{-}(\tau,T),T). \ \ Let \ k \in (k_{\min} \ ... \ k_{\max}] \ \ be such that \ \mathsf{R}_{f,s,k,H}^{-}(\tau,T) \neq \emptyset. \ \ Let \ also \ x = s + kp, \ c = \mathsf{weight\text{-}count}_{\mathcal{P}}(x,n), \ and \ c' \in [1 \ ... \ | \mathsf{R}_{f,s,k,H}^{-}(\tau,T)|].$ Then:

- It holds $c' \in [1...c]$, i.e., weight-select_P(x, n, c') is well-defined,
- Every position $j \in \text{weight-select}_{\mathcal{P}}(x, n, c')$ satisfies $j x \in \mathsf{R}^-_{f, s, k, H}(\tau, T)$

Proof. By Lemma 8.82, it holds $|\mathsf{R}_{f,s,k,H}^-(\tau,T)| = \mathsf{weight\text{-}count}_{\mathcal{P}}(s+kp,n) = \mathsf{weight\text{-}count}_{\mathcal{P}}(x,n) = c$. Thus, $c' \in [1 ... c]$, i.e., weight-select_P(x, n, c') is well-defined.

Let us now consider any $j \in \mathsf{weight\text{-}select}_{\mathcal{P}}(x, n, c')$. Then, by definition (see Section 6), there exists $(x', y', w', \ell') \in \mathcal{P}$ such that $x \leq x' < n$ and $j = \ell'$. Since $\mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f, H}^-(\tau, T), T)$, it follows by Definition 6.5 that there exists $(q,h) \in \operatorname{Seed}_{f,H}^-(\tau,T)$ such that x' = h and $j = \ell'$ satisfies $T^{\infty}[j - \ell']$ $7\tau ... j) = T^{\infty}[q - 7\tau ... q)$. Furthermore, by definition of Seed $_{f,H}^-(\tau,T)$ (see Section 8.4.4), there exists $q' \in$ $\operatorname{comp}_{14\tau}(\mathsf{R}_{f,H}^-(\tau,T),T)\subseteq\mathsf{R}_{f,H}^-(\tau,T)$ such that $q=e_f^{\operatorname{full}}(\tau,T,q')$ and $h=\min(e_f^{\operatorname{full}}(\tau,T,q')-b(\tau,T,q'),7\tau)$. Denote $q''=b(\tau,T,q')$. Putting everything together we thus obtain that $x\leq x'=h=\min(q-q'',7\tau)\leq q-q''$ and $T^{\infty}[j-7\tau ...j+7\tau)=T^{\infty}[q-7\tau ...q+7\tau)$. Note also that since by definition, it holds [q''...q'], it follows by $q' \in \mathsf{R}_{f,H}^-(\tau,T)$ and Lemma 8.48 that $q'' \in \mathsf{R}_{f,H}^-(\tau,T)$ and $e_f^{\mathrm{full}}(\tau,T,q'') = e_f^{\mathrm{full}}(\tau,T,q')$. Denote q''' = q - x. We will prove that $q''' \in \mathsf{R}_{f,s,k,H}^-(\tau,T)$. First, note that by the assumption $q - q'' \ge x$, we immediately obtain $q'' \le q'''$. On the other hand, $k > k_{\min}$ implies that $e(\tau,T,q'') - q''' \ge e_f^{\mathrm{full}}(\tau,T,q'') - q''' = e_f^{\mathrm{full}}(\tau,T,q') - q''' = q - q''' = x = s + kp \ge s + (k_{\min}+1) \cdot p = s + \lceil \frac{3\tau-1-s}{p} \rceil p \ge s + \frac{3\tau-1-s}{p} \cdot p \ge 3\tau-1$. Thus, it follows by $q''' \ge q''$ and Lemma 8.49(1), that $[q'' ... q'''] \subseteq \mathbb{R}(\tau, T)$. By Lemma 8.48, we thus obtain $q''' \in \mathbb{R}^-_{f,H}(\tau, T)$ and $e_f^{\mathrm{full}}(\tau, T, q''') = e_f^{\mathrm{full}}(\tau, T, q'') = e_f^{\mathrm{full}}(\tau, T, q') = q$. Thus, head $f(\tau, T, q''') = (e_f^{\mathrm{full}}(\tau, T, q''') - q''') \mod p = (q - q''') \mod p = x \mod p = (s + pk) \mod p = s$ and $\exp_f(\tau, T, q'') = \lfloor \frac{e_f^{\mathrm{full}}(\tau, T, q''') - q'''}{p} \rfloor = \lfloor \frac{q - q'''}{p} \rfloor = \lfloor \frac{s + pk}{p} \rfloor = k$. We have thus proved $q''' \in \mathbb{R}^-_{f,s,k,H}(\tau, T)$. Recall now that $T^{\infty}[j - 7\tau ...j + 7\tau) = T^{\infty}[q - 7\tau ...q + 7\tau)$. By $k \leq k_{\max}$, we have $x = s + kp \leq s + \lfloor \frac{7\tau - s}{p} \rfloor p \leq 7\tau$. Thus, we obtain $T^{\infty}[j - x ...j) = T^{\infty}[q - x ...q)$. On the other hand, by the assumption, we have $T^{\infty}[j ... j + \tau) = T^{\infty}[q ... q + \tau)$. Thus, letting $\ell' = x + \tau$, it holds $T^{\infty}[j - x ... j - x + \ell') = T^{\infty}[q - x ... q - x + \ell') = T^{\infty}[q''' ... q''' + \ell')$. Combining this with $j \in [1 ... n]$ and $e(\tau, T, q''') - q''' = (e(\tau, T, q''') - e_f^{\text{full}}(\tau, T, q''')) + (e_f^{\text{full}}(\tau, T, q''') - q''') < \tau + (e_f^{\text{full}}(\tau, T, q''') - q''') = e_f^{\text{full}}(\tau, T, q''') + e_f^{\text{full}}(\tau, T, q'''') + e_f^{\text{full}}(\tau, T, q''') + e_f^{\text{full}}(\tau, T,$ $\tau + (q - q''') = \tau + x = \ell'$, by Lemma 8.55 implies that $j - x \in R(\tau, T)$, head_f $(\tau, T, j - x) = \text{head}_f(\tau, T, q''')$, $\operatorname{root}_f(\tau, T, j - x) = \operatorname{root}_f(\tau, T, q'''), \operatorname{type}(f, \tau, j - x) = \operatorname{type}(f, \tau, q'''), \operatorname{and} \exp_f(\tau, T, j - x) = \exp_f(\tau, T, q''').$ Thus, $j - x \in \mathsf{R}^-_{f,s,k,H}(\tau,T)$.

Lemma 8.84. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Let $H \in \Sigma^+$, p = |H|, $s \in [0 ... p)$, $k_{\min} = \lceil \frac{3\tau - 1 - s}{p} \rceil - 1$, $k_{\max} = \lfloor \frac{7\tau - s}{p} \rfloor$, and $\mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f,H}^{-}(\tau,T),T)$. Let $k \in (k_{\min} ... k_{\max}]$ be such that $\mathsf{R}^-_{f,s,k,H}(\tau,T) \neq \emptyset$. Let also x=s+kp. Then, $c=\mathsf{weight\text{-}count}_{\mathcal{P}}(x,n)$ satisfies $c\geq 1$ and every position $j \in \text{weight-select}_{\mathcal{P}}(x, n, 1)$ satisfies $j - x \in \mathsf{R}^-_{f,s,k,H}(\tau, T)$.

Proof. The result follows by Lemma 8.83 with c'=1.

Lemma 8.85. Let $\tau \in [1..\lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Let $H \in \Sigma^+$ and $\mathcal{P} =$ IntStrPoints_{7 τ}(Seed_{f,H}(τ ,T),T) (Definition 6.5). For any integer $x_l \in [0..7\tau]$ and any $y_l, y_u \in \Sigma^*$ such $that \ \operatorname{weight-count}_{\mathcal{P}}(x_l,n,y_l,y_u) > 0, \ it \ holds \ \operatorname{r-min}_{\mathcal{P}}(x_l,n,y_l,y_u) = \min\{e_f^{\operatorname{full}}(\tau,T,j) : j \in \mathsf{R}_{f,H}^{\prime,-u}(\tau,T), x_l \leq t\}$ $e_f^{\mathrm{full}}(\tau, T, j) - j$, and $y_l \leq T^{\infty}[e_f^{\mathrm{full}}(\tau, T, j) \dots e_f^{\mathrm{full}}(\tau, T, j) + 7\tau) \prec y_u\}.$

Proof. Denote $\mathsf{R}_{\mathrm{comp}} = \mathrm{comp}_{14\tau}(\mathsf{R}_{f,H}^-(\tau,T),T)$, and recall that by Definition 8.60, we have $\mathrm{Seed}_{f,H}^-(\tau,T) = \{(e_f^{\mathrm{full}}(\tau,T,j), \min(e_f^{\mathrm{full}}(\tau,T,j) - b(\tau,T,j),7\tau)) : j \in \mathsf{R}_{\mathrm{comp}}\}$. The proof consists of two steps labeled (a)-(b). (a) Let Q and A be defined as in the proof of Lemma 8.81(1), i.e., $Q = \{T^{\infty}[e_f^{\text{full}}(\tau, T, i) - 7\tau ... e_f^{\text{full}}(\tau, T, i) + 7\tau) : i \in \mathsf{R}_{\text{comp}}, x_l \leq e_f^{\text{full}}(\tau, T, i) - b(\tau, T, i), \text{ and } y_l \leq T^{\infty}[e_f^{\text{full}}(\tau, T, i) ... e_f^{\text{full}}(\tau, T, i) + 7\tau) \prec y_u\}$ and $A = \{j \in [1 ... n] : T^{\infty}[j - 7\tau ... j + 7\tau) \in Q\}$. Denote $\mathcal{R} = \{(x, y, w, \ell) \in \mathcal{P} : x_l \leq x < n \text{ and } y_l \leq y \prec y_u\}$. In the first step, we prove that $\operatorname{r-min}_{\mathcal{P}}(x_l, n, y_l, y_u) = \min A$.

- Let $i = \min A$. By definition of A, we then have $T^{\infty}[i 7\tau ... i + 7\tau) \in Q$. Denote $S = T^{\infty}[i 7\tau ... i + 7\tau)$. Recall from the proof of Lemma 8.81(1), that the function $g: \mathcal{R} \to Q$ defined such that for every $p=(x,y,w,\ell)\in\mathcal{R},\ g(p)=T^{\infty}[\ell-7\tau..\ell+7\tau),$ is a bijection. Let us thus consider $p_S=g^{-1}(S)\in\mathcal{R}.$ Denote $p_S = (x_S, y_S, w_S, \ell_S)$. Observe, that we then have $T^{\infty}[\ell_S - 7\tau ...\ell_S + 7\tau) = g(p_S) = S =$ $T^{\infty}[i-7\tau ...i+7\tau)$. By Definition 6.5, it then holds $\ell_S = \min\{i' \in [1...n]: T^{\infty}[i'-7\tau ...i'+7\tau) = S\}$. Consequently, $\ell_S \leq i$ and thus $\operatorname{r-min}_{\mathcal{P}}(x_l, n, y_l, y_u) = \min_{(x, y, w, \ell) \in \mathcal{R}} \ell \leq \ell_S \leq i = \min A$.
- Let $p_{\min} = (x_{\min}, y_{\min}, w_{\min}, \ell_{\min}) \in \mathcal{R}$ be such that $\operatorname{r-min}_{\mathcal{P}}(x_l, n, y_l, y_u) = \min_{(x, y, w, \ell) \in \mathcal{R}} \ell = \ell_{\min}$. Let $S = g(p_{\min}) = T^{\infty}[\ell_{\min} - 7\tau ...\ell_{\min} + 7\tau)$, where $g : \mathcal{R} \to Q$ is the bijection defined above. We then have $S \in Q$. Consequently, by $\ell_{\min} \in [1 ... n]$ and $T^{\infty}[\ell_{\min} - 7\tau ... \ell_{\min} + 7\tau) = S$, we have $\ell_{\min} \in A$ Thus, $\min A \leq \ell_{\min} = \text{r-min}_{\mathcal{P}}(x_l, n, y_l, y_u)$.
- (b) Let A' be defined as in the proof of Lemma 8.81(1), i.e., $A' = \{e_f^{\text{full}}(\tau, T, j) : j \in \mathsf{R}_{f,H}'^{-}(\tau, T), x_l \leq 1\}$ (b) Let A be defined as in the proof of Lemma 6.67(1), Ref., $A' = \{e_f (t, T, j) : j \in \mathbb{N}_{f,H}(t, T), x_l \leq e_f^{\mathrm{full}}(\tau, T, j) - j$, and $y_l \leq T^{\infty}[e_f^{\mathrm{full}}(\tau, T, j) ... e_f^{\mathrm{full}}(\tau, T, j) + 7\tau) \prec y_u\}$. In the proof of Lemma 8.81(1), we showed that A = A'. Combining this with $r\text{-min}_{\mathcal{P}}(x_l, n, y_l, y_u) = \min A$ shown above, we obtain $r\text{-min}_{\mathcal{P}}(x_l, n, y_l, y_u) = \min A = \min A' = \min \{e_f^{\mathrm{full}}(\tau, T, j) : j \in \mathbb{N}_{f,H}'(\tau, T), x_l \leq e_f^{\mathrm{full}}(\tau, T, j) - j$, and $y_l \leq T^{\mathrm{full}}(\tau, T, j) = 0$. $T^{\infty}[e_f^{\text{full}}(\tau, T, j) \dots e_f^{\text{full}}(\tau, T, j) + 7\tau) \prec y_u\}$, i.e., the claim.

Lemma 8.86. Let $\tau \in [1 ... \lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Let $H \in \Sigma^+$, p = |H|, $s \in [0 ... p)$, $\mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f,H}^{-}(\tau,T),T), \ k_{\min} = \lceil \frac{3\tau - 1 - s}{p} \rceil - 1, \ and \ k_{\max} = \lfloor \frac{7\tau - s}{p} \rfloor. \ Let \ k \in (k_{\min} \ldots k_{\max}] \ be$ such that $R_{f,s,k,H}^-(\tau,T) \neq \emptyset$. Then:

- 1. It holds $\min \mathsf{R}^-_{f,s,k,H}(\tau,T) = \mathsf{r\text{-}min}_{\mathcal{P}}(s+kp,n) s kp$. 2. The position $j = \min \mathsf{R}^-_{f,s,k,H}(\tau,T)$ satisfies $j = \min \mathsf{Occ}_{4\ell}(j,T)$.

Proof. 1. For every $k \in (k_{\min} ... k_{\max}]$, denote $A_k := \{e_f^{\text{full}}(\tau, T, j) - s - kp : j \in \mathsf{R}'_{f,H}(\tau, T) \text{ and } s + kp \leq e_f^{\text{full}}(\tau, T, j) - j\}$ and $B_k := \{e_f^{\text{full}}(\tau, T, j) : j \in \mathsf{R}'_{f,H}(\tau, T) \text{ and } s + kp \leq e_f^{\text{full}}(\tau, T, j) - j\}$. Immediately from the definition, we have $\min A_k = \min B_k - s - kp$. By Lemma 8.58, we have $\mathsf{R}_{f,s,k,H}^-(\tau, T) = A_k$. Next, observe that by Lemma 8.82, we have $|\mathsf{R}_{f,s,k,H}^-(\tau,T)| = \mathsf{weight\text{-}count}_{\mathcal{P}}(s+kp,n)$. The assumption $\mathsf{R}^-_{f,s,k,H}(\tau,T) \neq \emptyset \text{ thus implies that weight-count}_{\mathcal{P}}(s+kp,n,\varepsilon,c^\infty) = \mathsf{weight-count}_{\mathcal{P}}(s+kp,n) > 0 \text{ (where } t = 0$ $c = \max \Sigma$). Consequently, by Lemma 8.85, it holds $\min B_k = \operatorname{r-min}_{\mathcal{P}}(s + kp, n, \varepsilon, c^{\infty}) = \operatorname{r-min}_{\mathcal{P}}(s + kp, n)$ (note that Lemma 8.85 requires that $s + kp \le 7\tau$, which holds here since $s + kp \le s + \lfloor \frac{7\tau - s}{p} \rfloor \cdot p \le 7\tau$). Putting everything together, we thus obtain $\min \mathsf{R}^-_{f,s,k,H}(\tau,T) = \min A_k = \min B_k - s - kp = \mathsf{r-min}_{\mathcal{P}}(s+kp,n) - s - kp$.

2. First, note that $j \in \mathsf{R}^-_{f,s,k,H}(\tau,T)$ implies that $e(\tau,T,j)-j=s+kp \leq s+\lfloor \frac{7\tau-s}{p} \rfloor p \leq 7\tau \leq 2\ell+\tau \leq 3\ell$. Suppose that $j \neq \min \mathsf{Occ}_{4\ell}(j,T)$. Then, there exists $j' \in [1\ldots j)$ such that $j' \in \mathsf{Occ}_{4\ell}(j,T)$, or equivalently (by $\text{Lemma 8.11) } T^{\infty}[j\mathinner{\ldotp\ldotp\ldotp} j+4\ell) = T^{\infty}[j'\mathinner{\ldotp\ldotp\ldotp} j'+4\ell). \text{ By Lemma 8.55 (with } \delta=0), \text{ we thus obtain } j'\in\mathsf{R}^-_{f,s,k,H}(\tau,T), j'+4\ell).$ contradicting the definition of j.

Weighted Modular Constraint Queries

Lemma 8.87. Let $\tau \in [1..\lfloor \frac{n}{2} \rfloor]$ and f be any necklace-consistent function. Let $H \in \Sigma^+$, p = |H|, $s \in [0..p)$, and $\mathcal{I} = \text{WeightedIntervals}_{7\tau}(\text{Seed}_{f,H}^-(\tau,T),T)$ (Definitions 7.1 and 8.60). Denote $k_{\min} = \lceil \frac{3\tau-1-s}{p} \rceil - 1$ and $k_{\max} = \lfloor \frac{7\tau-s}{p} \rfloor$. For every k_1 and k_2 such that $k_{\min} \leq k_1 \leq k_2 \leq k_{\max}$, it holds

$$\mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k_1,k_2) = |\{j \in \mathsf{R}^-_{f,s,H}(\tau,T) : \exp_f(\tau,T,j) \in (k_1 \mathinner{\ldotp\ldotp} k_2]\}|.$$

Proof. Denote $\mathsf{R}_{\mathrm{comp}} = \mathrm{comp}_{14\tau}(\mathsf{R}_{f,H}^-(\tau,T),T)$. Recall that by Definition 8.60, it holds $\mathrm{Seed}_{f,H}^-(\tau,T) = \{(e_f^{\mathrm{full}}(\tau,T,j),\min(7\tau,e_f^{\mathrm{full}}(\tau,T,j)-b(\tau,T,j))): j\in\mathsf{R}_{\mathrm{comp}}\}$. Let $q\in(k_{\min}..k_{\max}]$. The proof consists of four steps labeled (a) through (d).

(a) Denote

$$Q = \{ T^{\infty}[e_f^{\text{full}}(\tau, T, i) - 7\tau \dots e_f^{\text{full}}(\tau, T, i) + 7\tau) : i \in \mathsf{R}_{\text{comp}} \text{ and } s + pq \leq e_f^{\text{full}}(\tau, T, i) - b(\tau, T, i) \},$$

$$A = \{ j \in [1 \dots n] : T^{\infty}[j - 7\tau \dots j + 7\tau) \in Q \}.$$

In the first step, we prove that it holds $\operatorname{mod-count}_{\mathcal{I},p}(s,q-1,q) = |A|$. Let $\mathcal{I}' = \{(e,w,\ell) \in \mathcal{I} : s+pq \leq e\}$. We begin by proving that for every $u = (e_u,w_u,\ell_u) \in \mathcal{I}'$, it holds $T^{\infty}[\ell_u-7\tau ..\ell_u+7\tau) \in Q$. By $u \in \mathcal{I}$, there exists $i_u \in \mathsf{R}_{\operatorname{comp}}$ such that it holds $e_u = \min(7\tau,e_f^{\operatorname{full}}(\tau,T,i_u)-b(\tau,T,i_u))$, and ℓ_u satisfies $T^{\infty}[\ell_u-7\tau ..\ell_u+7\tau) = T^{\infty}[e_f^{\operatorname{full}}(\tau,T,i_u)-7\tau ..e_f^{\operatorname{full}}(\tau,T,i_u)+7\tau)$. By $u \in \mathcal{I}'$, we obtain $s+pq \leq e_u = \min(7\tau,e_f^{\operatorname{full}}(\tau,T,i_u)-b(\tau,T,i_u)-b(\tau,T,i_u)) \leq e_f^{\operatorname{full}}(\tau,T,i_u)-b(\tau,T,i_u)$. We thus have $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,i_u)-7\tau ..e_f^{\operatorname{full}}(\tau,T,i_u)+7\tau) \in Q$. By $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,i_u)-7\tau ..e_f^{\operatorname{full}}(\tau,T,i_u)+7\tau) = T^{\infty}[\ell_u-7\tau ..\ell_u+7\tau)$, we thus obtain $T^{\infty}[\ell_u-7\tau ..\ell_u+7\tau) \in Q$. Let $g:\mathcal{I}' \to Q$ be defined so that for every $u=(e,w,\ell) \in \mathcal{I}'$, we let $g(u)=T^{\infty}[\ell-7\tau ..\ell+7\tau)$. We prove that g is a bijection.

- Let $u_1, u_2 \in \mathcal{I}'$ be such that $g(u_1) = g(u_2)$. We show that $u_1 = u_2$. For $x \in \{1, 2\}$, denote $u_x = (e_x, w_x, \ell_x)$ and let $i_x \in \mathsf{R}_{\text{comp}}$ be such that $e_x = \min(7\tau, e_f^{\text{full}}(\tau, T, i_x) b(\tau, T, i_x))$ and $T^{\infty}[\ell_x 7\tau \dots \ell_x + 7\tau) = T^{\infty}[e_f^{\text{full}}(\tau, T, i_x) 7\tau \dots e_f^{\text{full}}(\tau, T, i_x) + 7\tau)$. Let also $S_x := T^{\infty}[\ell_x 7\tau \dots \ell_x + 7\tau) = T^{\infty}[e_f^{\text{full}}(\tau, T, i_x) 7\tau \dots e_f^{\text{full}}(\tau, T, i_x) + 7\tau)$. By the assumption, $S_1 = g(u_1) = g(u_2) = S_2$. This implies $\ell_1 = \min\{j \in [1 \dots n] : T^{\infty}[j 7\tau \dots j + 7\tau) = S_1\} = \min\{j \in [1 \dots n] : T^{\infty}[j 7\tau \dots j + 7\tau) = S_2\} = \ell_2$ and $w_1 = |\{j \in [1 \dots n] : T^{\infty}[j 7\tau \dots j 7\tau) = S_1\}| = |\{j \in [1 \dots n] : T^{\infty}[j 7\tau \dots j + 7\tau) = S_2\}| = w_2$. Observe that due to $T[n] = T^{\infty}[0]$ being unique in T, for every $j \in \mathsf{R}_{f,H}(\tau,T)$, letting $Z = T^{\infty}[e_f^{\text{full}}(\tau,T,j) 7\tau \dots e_f^{\text{full}}(\tau,T,j) + 7\tau)$, it holds $\min(7\tau, e_f^{\text{full}}(\tau,T,j) b(\tau,T,j)) = |H| + \log(Z[1 \dots 7\tau], Z[1 \dots 7\tau |H|]) = |H| + \log(S_2[1 \dots 7\tau], S_2[1 \dots 7\tau |H|]) = \min(7\tau, e_f^{\text{full}}(\tau,T,i_2) b(\tau,T,i_2)) = e_2$. Consequently, $u_1 = u_2$.
- $w_1 = |\{j \in [1 \dots n] : T^{\infty}[j-7\tau \dots j-7\tau) = S_1\}| = |\{j \in [1 \dots n] : T^{\infty}[j-7\tau \dots j+7\tau) = S_2\}| = w_2. \text{ Observe that due to } T[n] = T^{\infty}[0] \text{ being unique in } T, \text{ for every } j \in \mathsf{R}_{f,H}(\tau,T), \text{ letting } Z = T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j) 7\tau \dots e_f^{\mathrm{full}}(\tau,T,j) + 7\tau), \text{ it holds } \min(7\tau,e_f^{\mathrm{full}}(\tau,T,j) b(\tau,T,j)) = |H| + \mathrm{lcs}(Z[1 \dots 7\tau],Z[1 \dots 7\tau |H|]). \text{ Thus, } e_1 = \min(7\tau,e_f^{\mathrm{full}}(\tau,T,i_1) b(\tau,T,i_1)) = |H| + \mathrm{lcs}(S_1[1 \dots 7\tau],S_1[1 \dots 7\tau |H|]) = |H| + \mathrm{lcs}(S_2[1 \dots 7\tau],S_2[1 \dots 7\tau |H|]) = \min(7\tau,e_f^{\mathrm{full}}(\tau,T,i_2) b(\tau,T,i_2)) = e_2. \text{ Consequently, } u_1 = u_2.$ $\bullet \text{ Let } X \in Q. \text{ Then, there exits } i \in \mathsf{R}_{\mathrm{comp}} \text{ such that } s + pq \leq e_f^{\mathrm{full}}(\tau,T,i) b(\tau,T,i) \text{ and } T^{\infty}[e_f^{\mathrm{full}}(\tau,T,i) 7\tau \dots e_f^{\mathrm{full}}(\tau,T,i) + 7\tau) = X. \text{ Note that } i \in \mathsf{R}_{\mathrm{comp}} \text{ implies that we have } (e_f^{\mathrm{full}}(\tau,T,i), \min(7\tau,e_f^{\mathrm{full}}(\tau,T,i) b(\tau,T,i))) \in \mathrm{Seed}_{f,H}^-(\tau,T). \text{ Hence, there exists } u = (e,w,\ell) \in \mathcal{I} \text{ such that } e = \min(7\tau,e_f^{\mathrm{full}}(\tau,T,i) b(\tau,T,i)) b(\tau,T,i)) \text{ and } T^{\infty}[\ell 7\tau \dots \ell + 7\tau) = T^{\infty}[e_f^{\mathrm{full}}(\tau,T,i) 7\tau \dots e_f^{\mathrm{full}}(\tau,T,i) + 7\tau) = X. \text{ Observe now that } s + pq \leq s + pk_{\max} = s + p\lfloor \frac{7\tau s}{p} \rfloor \leq 7\tau \text{ and } s + pq \leq e_f^{\mathrm{full}}(\tau,T,i) b(\tau,T,i) \text{ imply that } s + pq \leq \min(7\tau,e_f^{\mathrm{full}}(\tau,T,i) b(\tau,T,i)). \text{ Thus, } s + pq \leq e_f^{\mathrm{consequently, it holds } (e,w,\ell) \in \mathcal{I}'. \text{ Since, as noted earlier, we have } T^{\infty}[\ell 7\tau \dots \ell + 7\tau) = X, \text{ it follows that } g(p) = T^{\infty}[\ell 7\tau \dots \ell + 7\tau) = X.$

We have thus proved that g is a bijection. Observe now that:

- $\bullet \text{ By definition, mod-count}_{\mathcal{I},p}(s,q-1,q) = \sum_{(e,w,\ell) \in \mathcal{I}} w \cdot |\{j \in [0\mathinner{\ldotp\ldotp\ldotp} e]: j \bmod p = s \text{ and } q-1 < \lfloor\frac{j}{p}\rfloor \leq q\}| = \sum_{(e,w,\ell) \in \mathcal{I}} w \cdot |\{j \in [0\mathinner{\ldotp\ldotp\ldotp} e]: j \bmod p = s \text{ and } \lfloor\frac{j}{p}\rfloor = q\}| = \sum_{(e,w,\ell) \in \mathcal{I}} w \cdot |\{j \in [0\mathinner{\ldotp\ldotp\ldotp\ldotp} e]: j = s+pq\}| = \sum_{(e,w,\ell) \in \mathcal{I}: s+pq \leq e} w = \sum_{(e,w,\ell) \in \mathcal{I}'} w.$
- $s+pq\}|=\sum_{(e,w,\ell)\in\mathcal{I}:s+pq\leq e} w=\sum_{(e,w,\ell)\in\mathcal{I}'} w.$ By Definition 7.1, we thus have $\operatorname{mod-count}_{\mathcal{I},p}(s,q-1,q)=\sum_{(e,w,\ell)\in\mathcal{I}'} w=\sum_{(e,w,\ell)\in\mathcal{I}'} |\{j\in[1..n]:T^{\infty}[j-7\tau\mathinner..j+7\tau)=T^{\infty}[\ell-7\tau\mathinner..\ell+7\tau)\}|$. Since g is a bijection, for any $(e_1,w_1,\ell_1),(e_2,w_2,\ell_2)\in\mathcal{I}',(e_1,w_1,\ell_1)\neq (e_2,w_2,\ell_2)$ implies $T^{\infty}[\ell_1-7\tau\mathinner..\ell_1+7\tau)\neq T^{\infty}[\ell_2-7\tau\mathinner..\ell_2+7\tau)$. Thus, in the above expression, no position $j\in[1\mathinner..n]$ is accounted twice (i.e., for different elements of \mathcal{I}'). On the other hand, g being a bijection also implies that for every $j\in[1\mathinner..n], T^{\infty}[j-7\tau\mathinner..j+7\tau)\in Q$ implies that there exists some $u=(e,w,\ell)\in\mathcal{I}'$ such that $T^{\infty}[j-7\tau\mathinner..j+7\tau)=T^{\infty}[\ell-7\tau\mathinner..\ell+7\tau)$. Consequently,

 $\begin{array}{l} \mathsf{mod\text{-}count}_{\mathcal{I},p}(s,q-1,q) = \sum_{(e,w,\ell) \in \mathcal{I}'} |\{j \in [1\mathinner{.\,.} n] : T^\infty[j-7\tau\mathinner{.\,.} j+7\tau) = T^\infty[\ell-7\tau\mathinner{.\,.} \ell+7\tau)\}| = \\ |\{j \in [1\mathinner{.\,.} n] : T^\infty[j-7\tau\mathinner{.\,.} j+7\tau) \in Q\}| = |A|. \end{array}$

(b) Denote

$$Q' = \{T^{\infty}[e_f^{\mathrm{full}}(\tau,T,i) - 7\tau \dots e_f^{\mathrm{full}}(\tau,T,i) + 7\tau) : i \in \mathsf{R}_{f,H}^-(\tau,T) \text{ and } s + pq \leq e_f^{\mathrm{full}}(\tau,T,i) - i\}.$$

In the second step, we prove that Q = Q'.

- Let $X \in Q$. Then, there exists $i \in \mathsf{R}_{\mathrm{comp}}$ such that $s + pq \leq e_f^{\mathrm{full}}(\tau, T, i) b(\tau, T, i)$ and $T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) 7\tau \dots e_f^{\mathrm{full}}(\tau, T, i) + 7\tau) = X$. By $\mathsf{R}_{\mathrm{comp}} \subseteq \mathsf{R}_{f,H}^{-}(\tau, T)$, we have $i \in \mathsf{R}_{f,H}^{-}(\tau, T)$. Denote $i' = b(\tau, T, i)$. By Lemma 8.48, we have $i' \in \mathsf{R}_{f,H}^{-}(\tau, T)$ and $e_f^{\mathrm{full}}(\tau, T, i') = e_f^{\mathrm{full}}(\tau, T, i)$. Consequently, $s + pq \leq e_f^{\mathrm{full}}(\tau, T, i) b(\tau, T, i) = e_f^{\mathrm{full}}(\tau, T, i') i'$. By definition of Q', $T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i') 7\tau \dots e_f^{\mathrm{full}}(\tau, T, i') + 7\tau) = Q'$. Recalling that $T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i') 7\tau \dots e_f^{\mathrm{full}}(\tau, T, i') + 7\tau) = T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) 7\tau \dots e_f^{\mathrm{full}}(\tau, T, i) + 7\tau) = T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) 7\tau \dots e_f^{\mathrm{full}}(\tau, T, i) + 7\tau) = T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) 7\tau \dots e_f^{\mathrm{full}}(\tau, T, i) + 7\tau) = T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) 7\tau \dots e_f^{\mathrm{full}}(\tau, T, i) 7\tau \dots e_f^{\mathrm{full}}(\tau, T, i) + 7\tau) = T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) 7\tau \dots e_f^{\mathrm{full}}(\tau, T, i) 7\tau \dots e_f^{\mathrm{full}}(\tau, T, i) + 7\tau) = T^{\infty}[e_f^{\mathrm{full}}(\tau, T, i) 7\tau \dots e_f^{\mathrm{full}}(\tau, T, i) 7\tau \dots e_f^{\mathrm{$ X, we thus obtain $X \in Q'$.
- Let $X \in Q'$. Then, there exists $i \in \mathsf{R}_{f,H}^-(\tau,T)$ such that $s+pq \leq e_f^{\mathrm{full}}(\tau,T,i)-i$ and $X = T^{\infty}[e_f^{\mathrm{full}}(\tau,T,i)-7\tau\ldots e_f^{\mathrm{full}}(\tau,T,i)+7\tau)$. Let $i' = \max(i,e_f^{\mathrm{full}}(\tau,T,i)-7\tau)$. Then, $[i\ldots i'] \subseteq \mathsf{R}(\tau,T)$. By Lemma 8.48, we thus have $i' \in \mathsf{R}_{f,H}^-(\tau,T)$ and $e_f^{\mathrm{full}}(\tau,T,i') = e_f^{\mathrm{full}}(\tau,T,i)$. By $s+pq \leq s+pk_{\max} = s+p\lfloor\frac{7\tau-s}{p}\rfloor \leq s+(7\tau-s)=7\tau$, we also have $e_f^{\mathrm{full}}(\tau,T,i')-i' \in [s+pq\ldots7\tau]$. Let $j=\min\{t\in[1\ldots n]:T^{\infty}[t-7\tau\ldots t+7\tau)=X\}$ and let j' be such that $e_f^{\mathrm{full}}(\tau,T,i')-i'=j-j'$. We prove the following properties of j':

 - First, as in the proof of Lemma 8.81, we obtain that $j' \in \mathbb{R}^-_{f,H}(\tau,T)$ and $e_f^{\mathrm{full}}(\tau,T,j') = j$. Second, we show that $s + pq \leq e_f^{\mathrm{full}}(\tau,T,j') j'$. First, recall that $e_f^{\mathrm{full}}(\tau,T,j') = j$. On the other hand, above we observed that $e_f^{\mathrm{full}}(\tau,T,i') i' \geq s + pq$. Thus, we obtain $s + pq \leq e_f^{\mathrm{full}}(\tau,T,i') i' = i$ $j - j' = e_f^{\text{full}}(\tau, T, j') - j'.$
 - Lastly, using the argument as in the proof of Lemma 8.81, we obtain $j' \in C(14\tau, T)$.

We have thus proved $j' \in \mathsf{R}^-_{f,H}(\tau,T) \cap \mathsf{C}(14\tau,T) = \mathsf{R}_{\text{comp}}$ and $s + pq \leq e_f^{\text{full}}(\tau,T,j') - j' \leq e_f^{\text{full}}(\tau,T,j') - b(\tau,T,j')$. Consequently, by definition of Q, $T^\infty[e_f^{\text{full}}(\tau,T,j') - 7\tau \dots e_f^{\text{full}}(\tau,T,j') + 7\tau) \in Q$. Therefore, by $T^\infty[e_f^{\text{full}}(\tau,T,j') - 7\tau \dots e_f^{\text{full}}(\tau,T,j') + 7\tau) = T^\infty[j - 7\tau \dots j + 7\tau) = X$, we have $X \in Q$.

(c) Denote

$$A' = \{ e_f^{\text{full}}(\tau, T, j) : j \in \mathsf{R}'_{f, H}(\tau, T) \text{ and } s + pq \le e_f^{\text{full}}(\tau, T, j) - j \}.$$

In the third step, we prove that A = A'.

- Let $j \in A$. Then, $j \in [1 ...n]$ and $T^{\infty}[j-7\tau...j+7\tau) \in Q'$ (recall, that Q=Q'). Thus, there exists $i \in \mathsf{R}_{f,H}^-(\tau,T)$ satisfying $s+pq \leq e_f^{\mathrm{full}}(\tau,T,i)-i$ and $T^{\infty}[e_f^{\mathrm{full}}(\tau,T,i)-7\tau...e_f^{\mathrm{full}}(\tau,T,i)+7\tau) = T^{\infty}[j-7\tau...j+7\tau)$. Let $i'=\max(i,e_f^{\mathrm{full}}(\tau,T,i)-7\tau)$. It holds $[i..i'] \subseteq \mathsf{R}(\tau,T)$ and hence by Lemma 8.48, we have $i' \in \mathsf{R}_{f,H}^-(\tau,T)$ and $e_f^{\mathrm{full}}(\tau,T,i') = e_f^{\mathrm{full}}(\tau,T,i)$. By $s+pq \leq 7\tau$, we also have $e_f^{\mathrm{full}}(\tau,T,i')-i' \in [s+pq...7\tau]$. Let j' be such that $e_f^{\mathrm{full}}(\tau,T,i')-i'=j-j'$. By the same argument as in the second step of this proof, we then have $j' \in \mathsf{R}_{f,H}^-(\tau,T)$ and $e_f^{\mathrm{full}}(\tau,T,j') = j$. Let $j''=b(\tau,T,j')$. By definition and Lemma 8.48, we have $j'' \in \mathsf{R}_{f,H}'(\tau,T)$ and $e_f^{\mathrm{full}}(\tau,T,j'')=e_f^{\mathrm{full}}(\tau,T,j')=j$. Observe now that by $j''=b(\tau,T,j')$, we have $j'' \leq j'$. Thus, $s+pq \leq e_f^{\mathrm{full}}(\tau,T,i')-i'=j-j' \leq j-j''=e_f^{\mathrm{full}}(\tau,T,j'')-j''$. We have thus proved that $j \in A'$ (with j'' as the position in $\mathsf{R}_{f,H}'(\tau,T)$ satisfying $e_f^{\mathrm{full}}(\tau,T,j'')=j$. $e_f^{\text{full}}(\tau, T, j'') = j).$
- $\begin{array}{l} e_f \quad (\tau, I, J) = J). \\ \bullet \quad \text{Let } j \in A'. \quad \text{Then, there exists } j' \in \mathsf{R}_{f,H}'^{-}(\tau,T) \text{ such that } s + pq \leq e_f^{\text{full}}(\tau,T,j') j' \text{ and } e_f^{\text{full}}(\tau,T,j') = j. \\ \text{To show } j \in A, \text{ we need to prove that } T^{\infty}[j 7\tau \mathinner{\ldotp\ldotp} j + 7\tau) \in Q' \text{ (recall that } Q = Q'), \text{ which in turn requires showing that there exists } t \in \mathsf{R}_{f,H}^{-}(\tau,T) \text{ such that } s + pq \leq e_f^{\text{full}}(\tau,T,t) t \text{ and } T^{\infty}[e_f^{\text{full}}(\tau,T,t) 7\tau \mathinner{\ldotp\ldotp\ldotp} e_f^{\text{full}}(\tau,T,t) + 7\tau) = T^{\infty}[j 7\tau \mathinner{\ldotp\ldotp\ldotp} j + 7\tau). \text{ Let } t = j'. \text{ Then, we indeed have } t \in \mathsf{R}_{f,H}^{-}(\tau,T) \subseteq \mathsf{R}_{f,H}^{-}(\tau,T), \ s + pq \leq e_f^{\text{full}}(\tau,T,t) t, \text{ and } T^{\infty}[e_f^{\text{full}}(\tau,T,t) 7\tau \mathinner{\ldotp\ldotp\cdotp} e_f^{\text{full}}(\tau,T,t) + 7\tau) = T^{\infty}[e_f^{\text{full}}(\tau,T,j') 7\tau \mathinner{\ldotp\ldotp\cdotp} e_f^{\text{full}}(\tau,T,j') + 7\tau) = T^{\infty}[j 7\tau \mathinner{\ldotp\ldotp\cdotp} j + 7\tau). \text{ Hence, } j \in A. \end{array}$
- (d) We now put everything together. By the first three steps, it holds $\mathsf{mod\text{-}count}_{\mathcal{I},p}(s,q-1,q) = |A| = |A'|$. On the other hand, since adding a fixed value to every element of a set does not change its cardinality, it

follows by Lemma 8.58(2), that $|A'| = |\mathbb{R}_{f,s,q,H}^-(\tau,T)|$. By Lemma 7.2(2), we thus have:

$$\begin{split} \operatorname{mod-count}_{\mathcal{I},p}(s,k_1,k_2) &= \sum_{q=k_1+1}^{k_2} \operatorname{mod-count}_{\mathcal{I},p}(s,q-1,q) \\ &= \sum_{q=k_1+1}^{k_2} |\mathsf{R}_{f,s,q,H}^-(\tau,T)| \\ &= |\{j \in \mathsf{R}_{f,s,H}^-(\tau,T) : \exp_f(\tau,T,j) \in (k_1 \mathinner{.\,.} k_2]\}|. \end{split}$$

Basic Navigation Primitives

Proposition 8.88. Let $k \in [4 .. \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Given CompSAPeriodic(T), the value k, and any $j \in R(\tau,T)$ such that $j = \min Occ_{2\ell}(j,T)$, we can in $\mathcal{O}(\log n)$ time compute head $f(\tau, T, j)$ and $|\operatorname{root}_f(\tau, T, j)|$.

Proof. Denote $R_{\text{comp}} := \text{comp}_{14\tau}(R(\tau, T), T) = R(\tau, T) \cap C(14\tau, T)$ (Definition 4.10). Observe that $14\tau \ge 2\ell$ (following from $\tau = \lfloor \frac{\ell}{3} \rfloor$ and $\ell \geq 16$). Thus, by Lemma 4.4, $j = \min \operatorname{Occ}_{2\ell}(j,T)$ implies that $j \in \mathsf{C}(14\tau,T)$. Thus, $j \in R_{\text{comp}}$. Consequently, there exists $i \in [1 ... n_{\text{runs},k}]$ (Section 8.2.2) such that, letting $A_{\text{runs},k}[i] =$ (p_i,t_i) , it holds $p_i \leq j < p_i + t_i$. Using binary search, we compute the index i in $\mathcal{O}(\log n)$ time. We then retrieve $A_{\text{root},k}[i] = (s,p) = (\text{head}_f(\tau,T,p_i),|\text{root}_f(\tau,T,p_i)|)$ (Section 8.4.4). Observe now that by $[p_i ... p_i + t_i) \subseteq \mathsf{R}(\tau, T)$ and the uniqueness of run-decomposition, the value $\mathsf{head}_f(\tau, T, p_i)$ determines $\operatorname{head}_f(\tau, T, p_i + \delta)$ for every $\delta \in [0 ... t_i)$. More precisely, $\operatorname{head}_f(\tau, T, p_i + \delta) = (s - \delta) \mod p$. Therefore, we have head $f(\tau, T, j) = (s - j + p_i) \mod p$. On the other hand, by Lemma 8.48, we have $|\operatorname{root}_f(\tau, T, j)| = |\operatorname{root}_f(\tau, T, j)|$ $|\operatorname{root}_f(\tau, T, p_i)| = p.$

Proposition 8.89. Let $k \in [4 .. \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, $f = f_{\tau,T}$ (Definition 8.64), and $j \in \mathsf{R}(\tau,T)$. Given CompSAPeriodic(T) and the values k, j, head_f(τ, T, j), and $|\operatorname{root}_f(\tau, T, j)|$, we can in $\mathcal{O}(\log n)$ time compute:

- $e(\tau, T, j)$,
- $\begin{array}{l} \bullet \ \exp_f(\tau,T,j), \\ \bullet \ e_f^{\mathrm{full}}(\tau,T,j), \\ \bullet \ e_f^{\mathrm{cut}}(\tau,T,j,\ell), \\ \bullet \ \exp_f^{\mathrm{cut}}(\tau,T,j,\ell). \end{array}$

Proof. Denote $s = \text{head}_f(\tau, T, j)$, $H = \text{root}_f(\tau, T, j)$, and p = |H|. Recall that using CompSACore(T) (which is part of CompSAPeriodic(T)), we can answer LCE_T queries in $\mathcal{O}(\log n)$ time. All values are thus consecutively computed as follows:

- $$\begin{split} \bullet & \ t := e(\tau,T,j) = j + p + \operatorname{LCE}_T(j,j+p), \\ \bullet & \ k := \exp_f(\tau,T,j) = \lfloor \frac{e(\tau,T,j) j \operatorname{head}_f(\tau,T,j)}{|\operatorname{root}_f(\tau,T,j)|} \rfloor = \lfloor \frac{t j s}{p} \rfloor, \\ \bullet & \ t' := e_f^{\operatorname{full}}(\tau,T,j) = j + \exp_f(\tau,T,j) \cdot |\operatorname{root}_f(\tau,T,j)| = j + kp, \\ \bullet & \ k_1 := \exp_f^{\operatorname{cut}}(\tau,T,j,\ell) = \min(\exp_f(\tau,T,j), \lfloor \frac{\ell \operatorname{head}_f(\tau,T,j)}{|\operatorname{root}_f(\tau,T,j)|} \rfloor) = \min(k, \lfloor \frac{\ell s}{p} \rfloor), \\ \bullet & \ k_2 := \exp_f^{\operatorname{cut}}(\tau,T,j,2\ell) = \min(k, \lfloor \frac{2\ell s}{p} \rfloor). \end{split}$$

In total, the query takes $\mathcal{O}(\log n)$ time.

Proposition 8.90. Let $k \in [4 ... \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Given CompSAPeriodic(T), the value k, and any $j \in R(\tau, T)$ such that $j = \min Occ_{2\ell}(j, T)$, we can in $O(\log n)$ time determine if $\operatorname{Seed}_{f,H}^{r}(\tau,T)\neq\emptyset$ (where $H=\operatorname{root}_{f}(\tau,T,j)$), and if so, return the pointer to the data structure from Proposition 6.6 for weighted range queries on IntStrPoints_{7 τ} (Seed⁻_{f,H}(τ , T), T) (Definitions 6.5 and 8.60).

Proof. First, recall that $2\ell \le 14\tau$. By Lemma 4.5 and Lemma 4.4, this implies that $j \in \text{comp}_{14\tau}(\mathbb{R}(\tau,T),T)$. Consequently, there exists $i \in [1..n_{\text{runs},k}]$ (Section 8.2.2) such that, letting $A_{\text{runs},k}[i] = (p_i, t_i)$, it holds $p_i \le j < p_i + t_i$. Using binary search over the array $A_{\text{runs},k}[1...n_{\text{runs},k}]$ (stored as part of CompSACore(T); see Section 8.2.2), we compute the index i in $\mathcal{O}(\log n)$ time. We then have $[p_i ... j] \subseteq \mathsf{R}(\tau, T)$, and hence

by Lemma 8.48, $\operatorname{root}_f(\tau, T, p_i) = \operatorname{root}_f(\tau, T, j) = H$. We then retrieve the pointer μ_H to the structure from Proposition 6.6 for Seed $_{t,H}^{-}(\tau,T)$ (i.e., answering range queries on IntStrPoints $_{7\tau}(\text{Seed}_{t,H}^{-}(\tau,T),T))$ from $A_{\text{ptr},k}[i]$ (Section 8.4.4). If μ_H is a null pointer (note that this is possible, since we did not assume $\operatorname{type}(\tau, T, j)$, then we return that $\operatorname{Seed}_{f,H}^-(\tau, T) = \emptyset$. Otherwise, we return μ_H as the answer.

Proposition 8.91. Let $k \in [4 .. \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Given CompSAPeriodic(T), the value k, and any $j \in R(\tau, T)$ such that $j = \min Occ_{2\ell}(j, T)$, we can in $O(\log n)$ time determine if $\operatorname{Seed}_{f,H}^-(\tau,T) \neq \emptyset$ (where $H = \operatorname{root}_f(\tau,T,j)$), and if so, return the pointer to the data structure from Proposition 7.3 for modular constraint queries on WeightedIntervals_{7 τ} (Seed $_{TH}^{\tau}(\tau,T),T$) (Definitions 7.1 and 8.60).

Proof. Similarly as in the proof of Proposition 8.90, using binary search over $A_{\text{runs},k}[1..n_{\text{runs},k}]$, we first compute $i \in [1..n_{\text{runs},k}]$ such that, letting $A_{\text{runs},k}[i] = (p_i,t_i)$, it holds $p_i \leq j < p_i + t_i$. We then have $[p_i \dots j] \subseteq \mathsf{R}(\tau,T)$, and hence by Lemma 8.48, $\mathsf{root}_f(\tau,T,p_i) = \mathsf{root}_f(\tau,T,j) = H$. We then retrieve the pointer μ_H to the structure from Proposition 7.3 for Seed $_{t,H}^-(\tau,T)$ (i.e., answering weighted modular constraint queries on WeightedIntervals_{7 τ} (Seed⁷_{f,H}(τ , T), T) from $A'_{\text{ptr},k}[i]$ (Section 8.4.4). If μ_H is a null pointer, we return that $\operatorname{Seed}_{f,H}^-(\tau,T)=\emptyset$. Otherwise, we return μ_H as the answer.

8.4.7 Computing the Size of Poslow and Poshigh

Combinatorial Properties

Lemma 8.92. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{type}(\tau, P) = -1$ and T[n] does not occur in P[1..m). Let $i \in R(\tau, T)$. Denote $H = \operatorname{root}_f(\tau, P)$ and $t = e(\tau, T, i) - i - 3\tau + 2$. Then, $|\operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P, T) \cap [i ... i + t)| \leq 1$ and $|\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \cap [i\mathinner{\ldotp\ldotp\ldotp} i+t)| \leq 1. \ \mathit{Moreover}, \ |\operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T) \cap [i\mathinner{\ldotp\ldotp\ldotp} i+t)| = 1 \ (\mathit{resp}. \ |\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \cap [i\mathinner{\ldotp\ldotp\ldotp} i+t)| = 1)$ holds if and only if

- $\begin{array}{l} \bullet \ i \in \mathsf{R}^-_{f,H}(\tau,T), \\ \bullet \ e^{\mathrm{full}}_f(\tau,T,i) i \geq e^{\mathrm{cut}}_f(\tau,P,\ell) 1 \ (resp.\ e^{\mathrm{full}}_f(\tau,T,i) i \geq e^{\mathrm{cut}}_f(\tau,P,2\ell) 1) \ and \\ \bullet \ P[e^{\mathrm{cut}}_f(\tau,P,\ell) \ldots \min(m,\ell)] \preceq T^\infty[e^{\mathrm{full}}_f(\tau,T,i) \ldots e^{\mathrm{full}}_f(\tau,T,i) + 7\tau) \\ (resp.\ P[e^{\mathrm{cut}}_f(\tau,P,2\ell) \ldots \min(m,2\ell)] \preceq T^\infty[e^{\mathrm{full}}_f(\tau,T,i) \ldots e^{\mathrm{full}}_f(\tau,T,i) + 7\tau)). \end{array}$

 $Lastly, \ if \ \operatorname{Pos}^{\operatorname{low}^-}_{f,\ell}(P,T) \cap [i\mathinner{\ldotp\ldotp\ldotp} i+t) \neq \emptyset \ (resp. \ \operatorname{Pos}^{\operatorname{high}^-}_{f,\ell}(P,T) \cap [i\mathinner{\ldotp\ldotp\ldotp} i+t) \neq \emptyset), \ then \ \operatorname{Pos}^{\operatorname{low}^-}_{f,\ell}(P,T) \cap [i\mathinner{\ldotp\ldotp\ldotp} i+t) = \{e^{\operatorname{full}}_f(\tau,T,i) - (e^{\operatorname{cut}}_f(\tau,P,\ell)-1)\} \ (resp. \ \operatorname{Pos}^{\operatorname{high}^-}_{f,\ell}(P,T) \cap [i\mathinner{\ldotp\ldotp\ldotp} i+t) = \{e^{\operatorname{full}}_f(\tau,T,i) - (e^{\operatorname{cut}}_f(\tau,P,2\ell)-1)\}).$

Proof. Below we only prove the lemma for $\operatorname{Pos_{f,\ell}^{low-}}(P,T)$. The version for $\operatorname{Pos_{f,\ell}^{high-}}(P,T)$ is identical, except we replace $e_f^{\mathrm{cut}}(\tau,P,\ell)$, $\exp_f^{\mathrm{cut}}(\tau,P,\ell)$, and ℓ with $e_f^{\mathrm{cut}}(\tau,P,2\ell)$, $\exp_f^{\mathrm{cut}}(\tau,P,2\ell)$, and 2ℓ , respectively. Note that below we also use the fact that $\ell \leq 7\tau$. Note that we also have $2\ell \leq 7\tau$.

Denote $s = \text{head}_f(\tau, P)$ and $k_1 = \exp_f^{\text{cut}}(\tau, P, \ell) = \min(\exp_f(\tau, P), \lfloor \frac{\ell - s}{|H|} \rfloor)$. First, note that by definition we have $e(\tau, T, i) - i \geq 3\tau - 1$. Thus, t > 0. By Lemma 8.49(1), it holds $[i \cdot e(\tau, T, i) - 3\tau + 1] = [i \cdot i + t) \subseteq [i \cdot i + t]$ R(τ, T). From Lemma 8.48, we thus obtain that for every $\delta \in [0..t)$, it holds $[i..e(\tau, T, i) \ \delta i + 1] = [i..t + t) \subseteq R(\tau, T)$. From Lemma 8.48, we thus obtain that for every $\delta \in [0..t)$, it holds $e_f^{\text{full}}(\tau, T, i + \delta) = e_f^{\text{full}}(\tau, T, i)$, which implies $e_f^{\text{full}}(\tau, T, i + \delta) - (i + \delta) = e_f^{\text{full}}(\tau, T, i) - i - \delta$. Consider any $j \in \text{Pos}_{f,\ell}^{\text{low}}(-P, T)$. By definition, we then have $j \in R_{f,s,k_1,H}^-(\tau, T)$. Thus, $e_f^{\text{full}}(\tau, T, j) - j = s + k_1 p = e_f^{\text{cut}}(\tau, P, \ell) - 1$. Consequently, $i + \delta \in \text{Pos}_{f,\ell}^{\text{low}}(-P, T)$ implies $e_f^{\text{full}}(\tau, T, i + \delta) - (i + \delta) = e_f^{\text{full}}(\tau, T, i) - i - \delta = e_f^{\text{cut}}(\tau, P, \ell) - 1$, i.e., $\delta = (e_f^{\text{full}}(\tau, T, i) - i) - (e_f^{\text{cut}}(\tau, P, \ell) - 1)$, and hence $|\text{Pos}_{f,\ell}^{\text{low}}(-P, T) \cap [i..i + t)| \leq 1$.

We now prove the equivalence. Let us first assume $|\text{Pos}_{f,\ell}^{\text{low}}(-P, T) \cap [i..i + t)| = 1$, i.e. that for some

We now prove the equivalence. Let us first assume $|Pos_{f,\ell}^{low}(P,T) \cap [i...i+t)| = 1$, i.e., that for some We now prove the equivalence. Let us first assume $|\operatorname{Pos}_{f,\ell}^{(i)}(P,T)| |[i\ldots i+t)| = 1$, i.e., that for some $\delta \in [0\ldots t)$, it holds $i+\delta \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T)$. As noted above, this implies $i+\delta \in \mathsf{R}_{f,s,k_1,H}^{-}(\tau,T)$. By $[i\ldots i+\delta]\subseteq \mathsf{R}(\tau,T)$ and Lemma 8.48, we thus have $i\in \mathsf{R}_{f,H}^{-}(\tau,T)$, i.e., the first condition. Next, recall from the above that $i+\delta \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T)$ implies $e_f^{\operatorname{cut}}(\tau,P,\ell)-1=e_f^{\operatorname{full}}(\tau,T,i)-i-\delta \leq e_f^{\operatorname{full}}(\tau,T,i)-i$. This establishes the second condition. We now show the third condition. By $k_1 \leq \exp_f(\tau,P)$, it holds $P[1\ldots e_f^{\operatorname{cut}}(\tau,P,\ell))=H'H^{k_1}$, where H' is a length-s suffix of H. On the other hand, by definition, $i+\delta \in \mathsf{R}_{f,s,k_1,H}^{-}(\tau,T)$ implies $T[i+\delta\ldots e_f^{\operatorname{full}}(\tau,T,i+\delta))=T[i+\delta\ldots e_f^{\operatorname{full}}(\tau,T,i))=H'H^{k_1}$. Therefore, the assumption $T[i+\delta\ldots n]\succeq P$ or $\operatorname{lcp}(P,T[i+\delta\ldots n])\geq \ell$ following from $i+\delta\in\operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T)$ is equivalent to $T[e_f^{\text{full}}(\tau,T,i)\ldots n]\succeq P[e_f^{\text{cut}}(\tau,P,\ell)\ldots m] \text{ or } \text{lcp}(P[e_f^{\text{cut}}(\tau,P,\ell)\ldots m],T[e_f^{\text{full}}(\tau,T,i)\ldots n]) \geq \ell - (e_f^{\text{cut}}(\tau,P,\ell)-1).$ Consider now two cases:

- First, assume $lcp(P[e_f^{cut}(\tau, P, \ell) \dots m], T[e_f^{full}(\tau, T, i) \dots n]) \ge \ell (e_f^{cut}(\tau, P, \ell) 1)$. Then, $m e_f^{cut}(\tau, P, \ell) + 1 \ge \ell (e_f^{cut}(\tau, P, \ell) 1)$, or equivalently $m \ge \ell$. By $\ell \le 7\tau$, we thus have $T^{\infty}[e_f^{full}(\tau, T, i) \dots e_f^{full}(\tau, T, i) + \ell (e_f^{cut}(\tau, P, \ell) 1)) = P[e_f^{cut}(\tau, P, \ell) \dots \ell] = P[e_f^{cut}(\tau, P, \ell) \dots \min(\ell, m)]$.
- $T = \{e_f(\tau, T, \ell) = 1, \text{ of equivalently } m \geq \ell. \text{ By } \ell \leq H, \text{ we this have } T = [e_f(\tau, T, \ell) \dots e_f(\tau, T, \ell) + H)$ $\geq T^{\infty}[e_f^{\text{till}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + \ell (e_f^{\text{cut}}(\tau, P, \ell) 1)) = P[e_f^{\text{cut}}(\tau, P, \ell) \dots \ell] = P[e_f^{\text{cut}}(\tau, P, \ell) \dots \min(\ell, m)].$ Let us now assume $T[e_f^{\text{full}}(\tau, T, i) \dots n] \geq P[e_f^{\text{cut}}(\tau, P, \ell) \dots m].$ This implies that for every $q \geq 0$, it holds $T^{\infty}[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + q] \geq P[e_f^{\text{cut}}(\tau, P, \ell) \dots \min(m, e_f^{\text{cut}}(\tau, P, \ell) + q)].$ In particular, for $q = 7\tau$, we have $T^{\infty}[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + 7\tau) \geq P[e_f^{\text{cut}}(\tau, P, \ell) \dots \min(m, e_f^{\text{cut}}(\tau, P, \ell) + 7\tau)] \geq P[e_f^{\text{cut}}(\tau, P, \ell) \dots \min(m, \ell)],$ where the last inequality follows by $e_f^{\text{cut}}(\tau, P, \ell) + 7\tau \geq \ell$.

To prove the opposite implication, consider any $i \in \mathsf{R}(\tau,T)$ and assume that it holds $i \in \mathsf{R}^-_{f,H}(\tau,T)$, $e^{\mathrm{full}}_f(\tau,T,i)-i \geq e^{\mathrm{cut}}_f(\tau,P,\ell)-1$, and $T^\infty[e^{\mathrm{full}}_f(\tau,T,i) \ldots e^{\mathrm{full}}_f(\tau,T,i)+7\tau) \succeq P[e^{\mathrm{cut}}_f(\tau,P,\ell) \ldots \min(m,\ell)]$. Let $\delta = (e^{\mathrm{full}}_f(\tau,T,i)-i)-(e^{\mathrm{cut}}_f(\tau,P,\ell)-1)$. We will prove that $\delta \in [0\ldots t)$ and $i+\delta \in \mathsf{Pos}^{\mathrm{low}}_{f,\ell}(P,T)$. The inequality $\delta \geq 0$ follows from the definition of δ and our assumptions. To show $\delta < t$, we consider two cases:

• First, let us assume that it holds $e(\tau,P)-1 < \ell$. Then, $k_1 = \exp^{\operatorname{cut}}_f(\tau,P,\ell) = \min(\exp_f(\tau,P),\lfloor\frac{\ell-s}{|H|}\rfloor) = \min(\lfloor\frac{e(\tau,P)-1-s}{|H|}\rfloor,\lfloor\frac{\ell-s}{|H|}\rfloor) = \lfloor\frac{e(\tau,P)-1-s}{|H|}\rfloor = \exp_f(\tau,P)$. Thus, $e^{\operatorname{cut}}_f(\tau,P,\ell) = 1+s+k_1|H|=1+s+\exp_f(\tau,P)|H| = e^{\operatorname{full}}_f(\tau,P)$. Denote $q:=\operatorname{tail}_f(\tau,P) = e(\tau,P)-e^{\operatorname{full}}_f(\tau,P)$ and $q':=\operatorname{tail}_f(\tau,T,i) = e(\tau,T,i)-e^{\operatorname{full}}_f(\tau,T,i)$. We show that $q' \geq q$. Suppose that q' < q. Note that by definition, $P[e^{\operatorname{full}}_f(\tau,P)...e(\tau,P)) = H[1...q]$ and $T[e^{\operatorname{full}}_f(\tau,T,i)...e(\tau,T,i)) = H[1...q']$. We therefore obtain that $\operatorname{lcp}(P[e^{\operatorname{full}}_f(\tau,P)...e(\tau,P)) = H[1...q]$ and $T[e^{\operatorname{full}}_f(\tau,T,i)...e(\tau,T,i)) = H[1...q']$. We therefore obtain that $\operatorname{lcp}(P[e^{\operatorname{full}}_f(\tau,P)...m],T[e^{\operatorname{full}}_f(\tau,T,i)...n]) \geq \min(q,q') = q'$. By the uniqueness of T[n] in T, we have $e(\tau,T,i) \leq n$. Thus, $e^{\operatorname{full}}_f(\tau,T,i) + q' \leq n$. Therefore, by $\operatorname{lcp}(P[e^{\operatorname{full}}_f(\tau,P)...m],T[e^{\operatorname{full}}_f(\tau,T,i)...n]) \geq q'$ and $\operatorname{type}(\tau,T,i) = -1$, we have $T[e^{\operatorname{full}}_f(\tau,T,i)+q'] = T[e(\tau,T,i)] \prec T[e(\tau,T,i)-|H|] = T[e^{\operatorname{full}}_f(\tau,T,i)+q'] + q' = H[1]$ and hence $T[e^{\operatorname{full}}_f(\tau,T,i)+q'] \prec T[e^{\operatorname{full}}_f(\tau,T,i)+q'] + q' = H[1] = P[e^{\operatorname{full}}_f(\tau,P)+q']$. Consequently, $T[e^{\operatorname{full}}_f(\tau,T,i)-e^{\operatorname{full}}_f(\tau,T,i)+q'] \prec P[e^{\operatorname{full}}_f(\tau,P)-e^{\operatorname{full}}_f(\tau,P)+q']$. Observe now that we have $e^{\operatorname{full}}_f(\tau,P)+q' \leq e^{\operatorname{full}}_f(\tau,P)+q-1=e(\tau,P)-1 \leq m$ and $e^{\operatorname{full}}_f(\tau,P)+q' \leq e^{\operatorname{full}}_f(\tau,P)+q' \leq e^{$

$$\begin{split} e(\tau,T,i) - (i+\delta) &= (e_f^{\text{full}}(\tau,T,i) - (i+\delta)) + (e(\tau,T,i) - e_f^{\text{full}}(\tau,T,i)) \\ &= (e_f^{\text{full}}(\tau,P) - 1) + (e(\tau,T,i) - e_f^{\text{full}}(\tau,T,i)) \\ &\geq (e_f^{\text{full}}(\tau,P) - 1) + (e(\tau,P) - e_f^{\text{full}}(\tau,P)) \\ &= e(\tau,P) - 1 \geq 3\tau - 1. \end{split}$$

• Let us now assume $e(\tau,P)-1\geq \ell$. Note that we then have $m\geq e(\tau,P)-1\geq \ell$. Thus, $\min(m,\ell)=\ell$. We then also have $k_1=\min(\exp_f(\tau,P),\lfloor\frac{\ell-s}{|H|}\rfloor)=\min(\lfloor\frac{e(\tau,P)-1-s}{|H|}\rfloor,\lfloor\frac{\ell-s}{|H|}\rfloor)=\lfloor\frac{\ell-s}{|H|}\rfloor$ and hence $e_f^{\mathrm{cut}}(\tau,P,\ell)\geq 1+s+\lfloor\frac{\ell-s}{|H|}\rfloor|H|>1+s+(\frac{\ell-s}{|H|}-1)|H|=\ell-(|H|-1).$ Denote $q:=1+\ell-e_f^{\mathrm{cut}}(\tau,P,\ell)$ and $q':= \mathrm{tail}_f(\tau,T,i)=e(\tau,T,i)-e_f^{\mathrm{tull}}(\tau,T,i).$ We prove that $q'\geq q$. Suppose q'< q. First, note that by the above, it holds q<|H|. By $e(\tau,P)-1\geq \ell$, we thus have $P[e_f^{\mathrm{cut}}(\tau,P,\ell)..\ell]=H[1..q].$ On the other hand, by definition we have $T[e_f^{\mathrm{full}}(\tau,T,i)..e(\tau,T,i))=H[1..q'].$ Thus, $\mathrm{lcp}(P[e_f^{\mathrm{cut}}(\tau,P,\ell)..m],T[e_f^{\mathrm{full}}(\tau,T,i)..n])\geq \min(q,q')=q'.$ By the uniqueness of T[n] in T, we have $e(\tau,T,i)\leq n.$ Thus, $e_f^{\mathrm{full}}(\tau,T,i)+q'\leq n.$ Therefore, by $\mathrm{lcp}(P[e_f^{\mathrm{cut}}(\tau,P,\ell)..m],T[e_f^{\mathrm{full}}(\tau,T,i)..n])\geq q'$ and $\mathrm{type}(\tau,T,i)=-1,$ we have $T[e_f^{\mathrm{full}}(\tau,T,i)+q']=T[e(\tau,T,i)]\prec T[e(\tau,T,i)-|H|]=T[e_f^{\mathrm{full}}(\tau,T,i)+q'-|H|],$ and hence $T[e_f^{\mathrm{full}}(\tau,T,i)+q']\prec T[e_f^{\mathrm{full}}(\tau,T,i)+q'-|H|]=P[e_f^{\mathrm{cut}}(\tau,P,\ell)+q'-|H|]=P[e_f^{\mathrm{cut}}(\tau,P,\ell)+q'-|H|]$ is well-defined since $e_f^{\mathrm{cut}}(\tau,P,\ell)>\ell-(|H|-1)\geq 2\tau$ and $|H|<\tau.$ Consequently, $T[e_f^{\mathrm{full}}(\tau,T,i)..e_f^{\mathrm{full}}(\tau,T,i)+q']\prec P[e_f^{\mathrm{cut}}(\tau,P,\ell)..e_f^{\mathrm{cut}}(\tau,P,\ell)+q'-|H|$. Observe now that we have $e_f^{\mathrm{cut}}(\tau,P,\ell)+q'\leq e_f^{\mathrm{cut}}(\tau,P,\ell)+q'-1=\ell.$ On the other hand,

 $e_f^{\mathrm{full}}(\tau,T,i) + q' < e_f^{\mathrm{full}}(\tau,T,i) + |\mathrm{root}_f(\tau,T,i)| < e_f^{\mathrm{full}}(\tau,T,i) + \tau < e_f^{\mathrm{full}}(\tau,T,i) + 7\tau. \text{ Thus, we obtain } T^{\infty}[e_f^{\mathrm{full}}(\tau,T,i) \dots e_f^{\mathrm{full}}(\tau,T,i) + 7\tau) \\ \prec P[e_f^{\mathrm{cut}}(\tau,P,\ell) \dots \ell] = P[e_f^{\mathrm{cut}}(\tau,P,\ell) \dots \ell], \text{ which contradicts } P[e_f^{\mathrm{cut}}(\tau,P,\ell) \dots \ell] = P[e_f^{\mathrm{cut}}(\tau,P,\ell) \dots \ell]$ our main assumption. We have thus proved $q' \geq q$, i.e., $e(\tau, T, i) - e_f^{\text{full}}(\tau, T, i) \geq 1 + \ell - e_f^{\text{cut}}(\tau, P, \ell)$.

$$\begin{split} e(\tau,T,i) - (i+\delta) &= (e_f^{\text{full}}(\tau,T,i) - (i+\delta)) + (e(\tau,T,i) - e_f^{\text{full}}(\tau,T,i)) \\ &= (e_f^{\text{cut}}(\tau,P,\ell) - 1) + (e(\tau,T,i) - e_f^{\text{full}}(\tau,T,i)) \\ &\geq (e_f^{\text{cut}}(\tau,P,\ell) - 1) + (1 + \ell - e_f^{\text{cut}}(\tau,P,\ell)) \\ &= \ell \geq 3\tau - 1. \end{split}$$

In both cases, we have shown $e(\tau, T, i) - (i + \delta) \ge 3\tau - 1$, or equivalently, $\delta \le e(\tau, T, i) - i - 3\tau + 1 < t$. In both cases, we have shown $e(\tau, I, i) - (i + \delta) \ge 3\tau - 1$, or equivalently, $\delta \le e(\tau, I, i) - i - 3\tau + 1 < t$. It remains to show $i + \delta \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T)$. First, recall that by $e_f^{\operatorname{full}}(\tau, T, i + \delta) = e_f^{\operatorname{full}}(\tau, T, i)$ and the definition of δ , we have $e_f^{\operatorname{full}}(\tau, T, i + \delta) - (i + \delta) = e_f^{\operatorname{full}}(\tau, T, i) - (i + \delta) = e_f^{\operatorname{cut}}(\tau, P, \ell) - 1$. Thus, since above we noted that $[i ... i + t) \subseteq \mathsf{R}_{f,H}^-(\tau,T)$ (and hence $i + \delta \in \mathsf{R}_{f,H}^-(\tau,T)$), we have $\operatorname{head}_f(\tau, T, i + \delta) = (e_f^{\operatorname{full}}(\tau, T, i + \delta) - (i + \delta)) \mod_f |H| = (e_f^{\operatorname{cut}}(\tau, P, \ell) - 1) \mod_f |H| = (\operatorname{head}_f(\tau, P) + k_1 |H|) \mod_f |H| = \operatorname{head}_f(\tau, P)$ and $\exp_f(\tau, T, i + \delta) = \lfloor \frac{e_f^{\operatorname{full}}(\tau, T, i + \delta) - (i + \delta)}{|H|} \rfloor = \lfloor \frac{e_f^{\operatorname{full}}(\tau, P, \ell) - 1}{|H|} \rfloor = k_1$. We have therefore proved that $i + \delta \in \mathsf{R}_{f,s,H}^-(\tau,T)$ and $\exp_f(\tau, T, i + \delta) = k_1$. Thus, to show $i + \delta \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T)$ it remains to prove that it holds $T^{\infty}[i + \delta ... n] \succeq P$ or $\operatorname{lcp}(P, T[i + \delta ... n]) \ge \ell$. To this end, first note that, letting H' be a length-s suffix of H by the above we have $T[i + \delta ... e^{\operatorname{full}}(\tau, T, i + \delta)] = T[i + \delta ... e^{\operatorname{full}}(\tau, T, i)) = P[1 - e^{\operatorname{cut}}(\tau, P, \ell)] = H'H^{k_1}$ H, by the above we have $T[i+\delta ...e_f^{\text{full}}(\tau,T,i+\delta)) = T[i+\delta ...e_f^{\text{full}}(\tau,T,i)) = P[1...e_f^{\text{cut}}(\tau,P,\ell)) = H'H^{k_1}$. Consequently, $lcp(P,T[i+\delta ...n]) = (e_f^{\text{cut}}(\tau,P,\ell)-1) + lcp(P[e_f^{\text{cut}}(\tau,P,\ell)...m],T[e_f^{\text{full}}(\tau,T,i)...n])$. Denote $h = lcp(P[e_f^{\text{cut}}(\tau,P,\ell)...m],T[e_f^{\text{full}}(\tau,T,i)...n])$. Consider two cases:

- First, assume $h \ge \ell (e_f^{\text{cut}}(\tau, P, \ell) 1)$. Then, we immediately obtain $\text{lcp}(P, T[i + \delta \dots n]) = (e_f^{\text{cut}}(\tau, P, \ell) 1)$. 1) $+h \ge \ell$, and hence $i + \delta \in \operatorname{Pos}_{f,\ell}^{\mathrm{low}^-}(P,T)$. • Second, assume $h < \ell - (e_f^{\mathrm{cut}}(\tau,P,\ell)-1)$. We will prove that $T[e_f^{\mathrm{full}}(\tau,T,i) \dots n] \succeq P[e_f^{\mathrm{cut}}(\tau,P,\ell) \dots m]$.
- Consider two subcases:
 - Let us first assume that $e_f^{\text{cut}}(\tau, P, \ell) + h \leq m$. Since T[n] does not occur in P[1..m), this implies $e_f^{\text{full}}(\tau, T, i) + h \leq n$. Thus, we obtain $T[e_f^{\text{full}}(\tau, T, i) + h] \neq P[e_f^{\text{cut}}(\tau, P, \ell) + h]$, and consequently $T[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + h] \neq P[e_f^{\text{cut}}(\tau, P, \ell) \dots e_f^{\text{cut}}(\tau, P, \ell) + h]$. We show that it holds $T[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + h] \succ P[e_f^{\text{cut}}(\tau, P, \ell) \dots e_f^{\text{cut}}(\tau, P, \ell) + h]$. To see this, assume $T[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + h] \prec P[e_f^{\text{cut}}(\tau, P, \ell) \dots e_f^{\text{cut}}(\tau, P, \ell) + h]$. Note that $h \leq \ell 1 \leq \tau 1$ and $e_f^{\text{cut}}(\tau, P, \ell) + h \leq \ell$. Thus, we also have $T^{\infty}[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + \tau 1] \prec P[e_f^{\text{cut}}(\tau, P, \ell) \dots \min(m, \ell)]$, which contradicts our assumption. Hence, $T[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + h] \succ P[e_f^{\text{cut}}(\tau, P, \ell) \dots e_f^{\text{cut}}(\tau, P, \ell) + h]$. This implies $T[e_f^{\text{full}}(\tau, T, i) \dots n] \succ P[e_f^{\text{cut}}(\tau, P, \ell) \dots m]$.

 Let us now assume $e_f^{\text{cut}}(\tau, P, \ell) + h = m + 1$. Since T[n] does not occur in $P[1 \dots m)$, the suffix $T[e_f^{\text{full}}(\tau, T, i) \dots n]$ can thus match with P only up to symbol T[n]. Thus, $e_f^{\text{full}}(\tau, T, i) + t = 0$.
 - suffix $T[e_f^{\text{full}}(\tau, T, i) ... n]$ can thus match with P only up to symbol T[n]. Thus, $e_f^{\text{full}}(\tau, T, i) + h 1 \le n$. Consequently, we have $T[e_f^{\text{full}}(\tau, T, i) ... n] \succeq T[e_f^{\text{full}}(\tau, T, i) ... e_f^{\text{full}}(\tau, T, i) + h 1] = P[e_f^{\text{cut}}(\tau, P, \ell) ... e_f^{\text{full}}(\tau, P, \ell) ... e_f^{\text{$

We have thus proved $T[e_f^{\mathrm{full}}(\tau, T, i) \dots n] \succeq P[e_f^{\mathrm{cut}}(\tau, P, \ell) \dots m]$. Combining with $T[i + \delta \dots e_f^{\mathrm{full}}(\tau, T, i)) = 0$ $P[1\mathinner{.\,.} e_f^{\mathrm{cut}}(\tau,P,\ell)), \text{ this implies } T[i+\delta\mathinner{.\,.} n] \succeq P. \text{ Thus, } i+\delta \in \mathrm{Pos}_{f,\ell}^{\mathrm{low}-}(P,T).$

To show the last implication, recall that we proved that $i+\delta\in\operatorname{Pos}_{f,\ell}^{\mathrm{low}-}(P,T)$ (where $\delta\in[0\mathinner{.\,.} t)$) implies $\delta=(e_f^{\mathrm{full}}(\tau,T,i)-i)-(e_f^{\mathrm{cut}}(\tau,P,\ell)-1)$. Thus, $\operatorname{Pos}_{f,\ell}^{\mathrm{low}-}(P,T)\cap[i\mathinner{.\,.} i+t)\neq\emptyset$ implies $\operatorname{Pos}_{f,\ell}^{\mathrm{low}-}(P,T)\cap[i\mathinner{.\,.} i+t)=\{i+\delta\}=\{e_f^{\mathrm{full}}(\tau,T,i)-(e_f^{\mathrm{cut}}(\tau,P,\ell)-1)\}.$

Lemma 8.93. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{type}(\tau, P) = -1$ and T[n] does not occur in P[1..m). Denote $H = \operatorname{root}_f(\tau, P)$, $s = \text{head}_f(\tau, P), k_1 = \exp_f^{\text{cut}}(\tau, P, \ell)$ (resp. $k_2 = \exp_f^{\text{cut}}(\tau, P, 2\ell)$). For every $i \in \mathsf{R}(\tau, T), i \in \mathsf{Pos}_{f,\ell}^{\mathrm{low}}(P, T)$ (resp. $i \in Pos_{f,\ell}^{high-}(P,T)$) holds if and only if

Proof. Below we only prove the lemma for $\operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T)$. The version for $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$ is identical, except we replace $\operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T)$, k_1 , $e_f^{\operatorname{cut}}(\tau,P,\ell)$, and ℓ with $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$, k_2 , $e_f^{\operatorname{cut}}(\tau,P,2\ell)$, and 2ℓ , respectively. Let $i \in \mathsf{R}(\tau,T)$ and assume $i \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T)$. By definition of $\operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T)$, it then holds $i \in \mathsf{R}_{f,s,H}^-(\tau,T)$ and $\operatorname{exp}_{f,\ell}(\tau,T) = k_1$, i.e. $i \in \mathsf{R}^-$

and $\exp_f(\tau, T, i) = k_1$, i.e., $i \in \mathsf{R}^-_{f,s,k_1,H}(\tau,T)$. This proves the first claim. To show the second claim, note that by $e(\tau, T, i) - i \geq 3\tau - 1$, letting $t = e(\tau, T, i) - i - 3\tau + 2$, we also have $\mathsf{Pos}^{\mathsf{low}}_{f,\ell}(P,T) \cap [i \dots i + t) \neq \emptyset$. By Lemma 8.92, we thus obtain $P[e_f^{\mathsf{cut}}(\tau, P, \ell) \dots \min(m, \ell)] \leq T^{\infty}[e_f^{\mathsf{full}}(\tau, T, i) \dots e_f^{\mathsf{full}}(\tau, T, i) + 7\tau)$.

Let us consider $i \in \mathsf{R}(\tau,T)$ and assume that it holds $i \in \mathsf{R}_{f,s,k_1,H}^{-}(\tau,T)$ and $P[e_f^{\mathrm{cut}}(\tau,P,\ell) \dots \min(m,\ell)] \leq T^{\infty}[e_f^{\mathrm{full}}(\tau,T,i) \dots e_f^{\mathrm{full}}(\tau,T,i) + 7\tau)$. The first assumption implies that $e_f^{\mathrm{full}}(\tau,T,i) - i = \mathrm{head}_f(\tau,T,i) + 7\tau$. $\exp_f(\tau, T, i) \cdot |\operatorname{root}_f(\tau, T, i)| = s + k_1 \cdot |H| = \operatorname{head}_f(\tau, P) + \exp_f^{\operatorname{cut}}(\tau, P, \ell) \cdot |\operatorname{root}_f(\tau, P)| = e_f^{\operatorname{cut}}(\tau, P, \ell) - 1 = e_f^{\operatorname{cut}}(\tau, P, \ell) - 1.$ By Lemma 8.92, it thus follows that, letting $t = e(\tau, T, i) - i - 3\tau + 2 > 0$, we have $\begin{array}{l} \operatorname{Pos}_{f,\ell}^{\mathrm{low}-}(P,T) \cap [i\mathinner{\ldotp\ldotp\ldotp} i+t) = \{e_f^{\mathrm{full}}(\tau,T,i) - (e_f^{\mathrm{cut}}(\tau,P,\ell)-1)\} = \{e_f^{\mathrm{full}}(\tau,T,i) - (e_f^{\mathrm{full}}(\tau,T,i)-i)\} = \{i\}. \ \, \text{Thus,} \\ i \in \operatorname{Pos}_{f,\ell}^{\mathrm{low}-}(P,T). \end{array}$

Lemma 8.94. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{type}(\tau, P) = -1$ and T[n] does not occur in P[1..m). Let $x_l = e_f^{\operatorname{cut}}(\tau, P, \ell) - 1$, $y_u = P[e_f^{\text{cut}}(\tau, P, \ell) ... \min(m, \ell)], \ x_l' = e_f^{\text{cut}}(\tau, P, 2\ell) - 1, \ and \ y_u' = P[e_f^{\text{cut}}(\tau, P, 2\ell) ... \min(m, 2\ell)]. \ Then, \ letting$ $H = \operatorname{root}_f(\tau, P)$, it holds:

- $\operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T) = \{e_f^{\operatorname{full}}(\tau,T,j) x_l : j \in \mathsf{R}_{f,H}^{\prime-}(\tau,T), \ x_l \leq e_f^{\operatorname{full}}(\tau,T,j) j, \ \text{and} \ y_u \preceq T^{\infty}[e_f^{\operatorname{full}}(\tau,T,j) \cdot e_f^{\operatorname{full}}(\tau,T,j) + 7\tau)\},$ $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) = \{e_f^{\operatorname{full}}(\tau,T,j) x_l' : j \in \mathsf{R}_{f,H}^{\prime-}(\tau,T), \ x_l' \leq e_f^{\operatorname{full}}(\tau,T,j) j, \ \text{and} \ y_u' \preceq T^{\infty}[e_f^{\operatorname{full}}(\tau,T,j) \cdot e_f^{\operatorname{full}}(\tau,T,j) + 7\tau)\}.$

Proof. Below we prove only the first formula. The proof for the second formula is analogous. By Proof. Below we prove only the first formula. The proof for the second formula is analogous. By Definition 8.70, we have $\operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T) \subseteq \mathsf{R}_{f,H}^{-}(\tau,T)$. On the other hand, by Lemma 8.49(1) it holds $\mathsf{R}_{f,H}^{-}(\tau,T) = \bigcup_{j \in \mathsf{R}_{f,H}^{\prime}(\tau,T)}[j \ldots e(\tau,T,j) - 3\tau + 1]$. Since this is a disjoint union, we thus have $\operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T) = \bigcup_{j \in \mathsf{R}_{f,H}^{\prime}(\tau,T)} \operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T) \cap [j \ldots e(\tau,T,j) - 3\tau + 1]$. By Lemma 8.92, for every $j \in \mathsf{R}_{f,H}^{\prime}(\tau,T)$, $|\operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T) \cap [j \ldots e(\tau,T,j) - 3\tau + 1]| = 1$ holds if and only if $x_l \leq e^{\operatorname{full}}(\tau,T,j) - j$ and $y_u \leq T^{\infty}[e^{\operatorname{full}}_f(\tau,T,j) \ldots e^{\operatorname{full}}_f(\tau,T,j) + 7\tau)$. Furthermore, if $x_l \leq e^{\operatorname{full}}_f(\tau,T,j) - j$ and $y_u \leq T^{\infty}[e^{\operatorname{full}}_f(\tau,T,j) - j - j]$. Consequently, letting $\mathcal{J} = \{j \in \mathsf{R}_{f,H}^{\prime}(\tau,T) : x_l \leq e^{\operatorname{full}}_f(\tau,T,j) - j \text{ and } y_u \leq T^{\infty}[e^{\operatorname{full}}_f(\tau,T,j) \ldots e^{\operatorname{full}}_f(\tau,T,j) + 7\tau)\}$ we thus obtain $\operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T) = \bigcup_{j \in \mathsf{R}_{f,H}^{\prime}(\tau,T)} \operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T) \cap [j \ldots e(\tau,T,j) - 3\tau + 1] = \bigcup_{j \in \mathcal{J}} \{e^{\operatorname{full}}_f(\tau,T,j) - x_l\}$, i.e., the claim. i.e., the claim.

Lemma 8.95. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that type $(\tau, P) = -1$ and T[n] does not occur in P[1..m). Let $x_l = e_f^{\text{cut}}(\tau, P, \ell) - 1$, $y_u = P[e_f^{\text{cut}}(\tau, P, \ell) ... \min(m, \ell)], \ x_l' = e_f^{\text{cut}}(\tau, P, 2\ell) - 1, \ and \ y_u' = P[e_f^{\text{cut}}(\tau, P, 2\ell) ... \min(m, 2\ell)]. \ Then, \ letting$ $H = \operatorname{root}_{f}(\tau, P)$ and $\mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{fH}^{-}(\tau, T), T)$, it holds:

- $$\begin{split} \bullet & |\operatorname{Pos}^{\mathrm{low}-}_{f,\ell}(P,T)| = \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l,n) \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l,n,y_u), \\ \bullet & |\operatorname{Pos}^{\mathrm{high}-}_{f,\ell}(P,T)| = \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l',n) \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l',n,y_u'). \end{split}$$

Proof. By Lemmas 8.81 and 8.94, it holds

$$\begin{split} |\operatorname{Pos}^{\operatorname{low}-}_{f,\ell}(P,T)| &= |\{j' \in \mathsf{R}'^-_{f,H}(\tau,T) : x_l \leq e_f^{\operatorname{full}}(\tau,T,j') - j' \text{ and } y_u \preceq T^{\infty}[e_f^{\operatorname{full}}(\tau,T,j') \ldots e_f^{\operatorname{full}}(\tau,T,j') + 7\tau)\}| \\ &= |\{j' \in \mathsf{R}'^-_{f,H}(\tau,T) : x_l \leq e_f^{\operatorname{full}}(\tau,T,j') - j'\}| - \\ &\quad |\{j' \in \mathsf{R}'^-_{f,H}(\tau,T) : x_l \leq e_f^{\operatorname{full}}(\tau,T,j') - j' \text{ and } T^{\infty}[e_f^{\operatorname{full}}(\tau,T,j') \ldots e_f^{\operatorname{full}}(\tau,T,j') + 7\tau) \prec y_u\}| \\ &= \operatorname{weight-count}_{\mathcal{P}}(x_l,n) - \operatorname{weight-count}_{\mathcal{P}}(x_l,n,y_u). \end{split}$$

Note that Lemma 8.81 requires that $x_l, x_l' \in [0..7\tau]$, which holds here since $x_l = e_f^{\text{cut}}(\tau, P, \ell) - 1 \le \ell$, and for $\tau = \lfloor \frac{\ell}{3} \rfloor$ and $\ell \geq 16$, it holds $\ell \leq 7\tau$. The proof for $|\operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T)|$ is analogous, except we use that $x_l' = e_f^{\text{cut}}(\tau, P, 2\ell) - j \le 2\ell$, and for $\tau = \lfloor \frac{\ell}{3} \rfloor$ and $\ell \ge 16$, it holds $2\ell \le 7\tau$.

Lemma 8.96. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ periodic pattern such that $\operatorname{type}(\tau, P) = -1$, $e(\tau, P) \leq |P|$, and T[n] does not occur in P[1..m). Let also $j_1 \in$ $\begin{aligned} &\operatorname{Occ}_{\ell}(P,T) \ and \ j_{2} \in \operatorname{Occ}_{2\ell}(P,T). \ Then, \ it \ holds \ j_{1}, j_{2} \in \mathsf{R}(\tau,T). \ Moreover, \ letting \ x_{l} = e_{f}^{\operatorname{cut}}(\tau,T,j_{1},\ell) - j_{1}, \\ &y_{u} = T[e_{f}^{\operatorname{cut}}(\tau,T,j_{1},\ell) \ldots j_{1} + \min(m,\ell)), \ x'_{l} = e_{f}^{\operatorname{cut}}(\tau,T,j_{2},2\ell) - j_{2}, \ y'_{u} = T[e_{f}^{\operatorname{cut}}(\tau,T,j_{2},2\ell) \ldots j_{2} + \min(m,2\ell)), \\ &H = \operatorname{root}_{f}(\tau,P), \ and \ \mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f,H}^{-}(\tau,T),T), \ it \ holds \end{aligned}$

- $$\begin{split} \bullet & |\operatorname{Pos}^{\operatorname{low}^-}_{f,\ell}(P,T)| = \operatorname{weight-count}_{\mathcal{P}}(x_l,n) \operatorname{weight-count}_{\mathcal{P}}(x_l,n,y_u), \\ \bullet & |\operatorname{Pos}^{\operatorname{high}^-}_{f,\ell}(P,T)| = \operatorname{weight-count}_{\mathcal{P}}(x_l',n) \operatorname{weight-count}_{\mathcal{P}}(x_l',n,y_u'). \end{split}$$

Proof. We show the claim only for $\operatorname{Pos}_{f,\ell}^{\operatorname{low}-}(P,T)$ (the proof for $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$ is analogous). By $\ell \geq 3\tau - 1$ and Lemma 8.50(1), it holds $j_1 \in \mathsf{R}(\tau,T)$, head $j_1 \in \mathsf{R}(\tau,T)$, head $j_1 \in \mathsf{R}(\tau,T)$ and $j_1 \in \mathsf{R}(\tau,T)$ by Lemma 8.56(1), we moreover have $\exp_{j_1}^{\operatorname{cut}}(\tau,T,j_1,\ell) = \exp_{j_1}^{\operatorname{cut}}(\tau,P,\ell)$. This implies that $e_{j_1}^{\operatorname{cut}}(\tau,P,\ell) = \exp_{j_1}^{\operatorname{cut}}(\tau,P,\ell)$. $1 = \operatorname{head}_{f}(\tau, P) + \exp_{f}^{\operatorname{cut}}(\tau, P, \ell) \cdot |\operatorname{root}_{f}(\tau, P)| = \operatorname{head}_{f}(\tau, T, j_{1}) + \exp_{f}^{\operatorname{cut}}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1})| = \operatorname{head}_{f}(\tau, T, j_{1}) + \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1})| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell) \cdot |\operatorname{root}_{f}(\tau, T, j_{1}, \ell)| = \operatorname{head}_{f}(\tau, T, j_{1}, \ell)|$ $e_f^{\mathrm{cut}}(\tau,T,j_1,\ell)-j_1=x_l$. By definition of $\mathrm{Occ}_\ell(P,T)$, we have $\mathrm{lcp}(T[j_1\ldots n],P)\geq \min(m,\ell)$. Consequently, $P[e_f^{\text{cut}}(\tau, P, \ell) ... \min(m, \ell)] = T[j_1 + e_f^{\text{cut}}(\tau, P, \ell) - 1 ... j_1 + \min(m, \ell) - 1] = T[e_f^{\text{cut}}(\tau, T, j_1, \ell) ... j_1 + \min(m, \ell)) = y_u$. Consequently, it follows by Lemma 8.95 that $|\operatorname{Pos}_{f,\ell}^{\text{low}-}(P, T)| = \operatorname{weight-count}_{\mathcal{P}}(x_l, n) - 1$ weight-count_{\mathcal{P}} (x_l, n, y_u) .

Lemma 8.97. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $j \in \mathsf{R}(\tau,T)$. Denote $s = \operatorname{head}_{f}(\tau, T, j), \ H = \operatorname{root}_{f}(\tau, T, j), \ p = |H|, \ H' = H(p - s ... p]. \ \text{Let P be a length-} 2\ell \ \operatorname{prefix of } H'H^{\infty}.$ $Then, \ P \ is \ \tau\text{-periodic. Moreover, letting } k_{1} = \lfloor \frac{\ell - s}{p} \rfloor, \ k_{2} = \lfloor \frac{2\ell - s}{p} \rfloor, \ x_{l} = s + k_{1}p, \ y_{u} = T[j + s ... j + s + \ell - x_{l}),$ $x'_{l} = s + k_{2}p, \ y'_{u} = T[j + s ... j + s + 2\ell - x'_{l}), \ \text{and P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f,H}^{-}(\tau, T), T), \ \text{it holds:}$

- $$\begin{split} \bullet & | \mathrm{Pos}^{\mathrm{low}-}_{f,\ell}(P,T) | = \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l,n) \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l,n,y_u), \\ \bullet & | \mathrm{Pos}^{\mathrm{high}-}_{f,\ell}(P,T) | = \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l',n) \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l',n,y_u'). \end{split}$$

Proof. By Lemma 8.57, P is τ -periodic, root_f $(\tau, P) = H$, type $(\tau, P) = -1$, T[n] does not occur in $P[1...2\ell)$, $e_f^{\text{cut}}(\tau, P, \ell) - 1 = x_l, \ e_f^{\text{cut}}(\tau, P, 2\ell) = x_l', \ P[e_f^{\text{cut}}(\tau, P, \ell) \dots \ell] = y_u, \ \text{and} \ P[e_f^{\text{cut}}(\tau, P, 2\ell) \dots 2\ell] = y_u'.$ The claims therefore follow by Lemma 8.95.

Lemma 8.98. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $j \in \mathbb{R}^-(\tau, T)$, $j_1 \in \mathbb{R}^+(\tau, T)$ $\begin{aligned} &\operatorname{Occ}_{\ell}(j,T), \ and \ j_{2} \in \operatorname{Occ}_{2\ell}(j,T). \ Then, \ it \ holds \ j_{1}, j_{2} \in \mathsf{R}(\tau,T). \ Moreover, \ letting \ x_{l} = e_{f}^{\operatorname{cut}}(\tau,T,j_{1},\ell) - j_{1}, \\ &y_{u} = T[e_{f}^{\operatorname{cut}}(\tau,T,j_{1},\ell) \ldots \min(n+1,j_{1}+\ell)), \ x'_{l} = e_{f}^{\operatorname{cut}}(\tau,T,j_{2},2\ell) - j_{2}, \ y'_{u} = T[e_{f}^{\operatorname{cut}}(\tau,T,j_{2},2\ell) \ldots \min(n+1,j_{2}+2\ell)), \ H = \operatorname{root}_{f}(\tau,T,j), \ and \ \mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{\overline{f},H}(\tau,T),T), \ it \ holds \end{aligned}$

- $$\begin{split} \bullet & |\operatorname{Pos}^{\mathrm{low}-}_{f,\ell}(j,T)| = \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l,n) \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l,n,y_u), \\ \bullet & |\operatorname{Pos}^{\mathrm{high}-}_{f,\ell}(j,T)| = \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l',n) \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l',n,y_u'). \end{split}$$

Proof. Denote P:=T[j..n] and m:=|P|=n-j+1. By definition, $j\in \mathbb{R}^-(\tau,T)$ implies that P is τ -periodic, type $(\tau, P) = -1$, $e(\tau, P) \leq |P|$, and that T[n] does not occur in P[1..m). We then also have $j_1 \in \operatorname{Occ}_{\ell}(j,T) = \operatorname{Occ}_{\ell}(T[j\mathinner{\ldotp\ldotp\ldotp} n],T) = \operatorname{Occ}(P,T) \text{ and } j_2 \in \operatorname{Occ}_{2\ell}(j,T) = \operatorname{Occ}_{2\ell}(T[j\mathinner{\ldotp\ldotp\ldotp} n],T) = \operatorname{Occ}_{2\ell}(P,T).$ Next, observe that $j_1 + \min(m, \ell) = \min(n+1, j_1 + \ell)$. To see this, consider two cases:

- First, assume $m < \ell$. The assumption $j_1 \in \mathrm{Occ}_{\ell}(j,T)$ implies $\mathrm{lcp}(T[j_1 \dots n],P) \geq \min(m,\ell) = m$. Thus, $T[j_1...j_1+m)=P$. By P[m]=T[n] and the uniqueness of T[n] in T, we then have $j_1+m-1=n$. Consequently, $j_1 + \min(m, \ell) = j_1 + \min(n - j_1 + 1, \ell) = \min(n + 1, j_1 + \ell)$.
- Second, assume $m \geq \ell$. The assumption $j_1 \in \text{Occ}_{\ell}(j,T)$ then implies $\text{lcp}(T[j_1 \dots n], P) \geq \min(m,\ell) = \ell$. Thus, $T[j_1 \dots j_1 + \ell) = P[1 \dots \ell]$. This implies that $\min(n+1, j_1 + \ell) = j_1 + \ell$, and hence $j_1 + \min(m, \ell) = j_1 + \ell$. $j_1 + \ell = \min(n+1, j_1 + \ell).$

We thus have $T[e_f^{\text{cut}}(\tau, T, j_1, \ell) \dots j_1 + \min(m, \ell)) = T[e_f^{\text{cut}}(\tau, T, j_1, \ell) \dots \min(n+1, j_1+\ell)) = y_u$. Consequently, the claim follows by Lemma 8.96.

Lemma 8.99. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $j \in \mathsf{R}^-(\tau,T)$, $x_l = e_f^{\text{cut}}(\tau, T, j, \ell) - j, \ y_u = T[e_f^{\text{cut}}(\tau, T, j, \ell) \dots \min(n+1, j+\ell)), \ x_l' = e_f^{\text{cut}}(\tau, T, j, 2\ell) - j, \ and \ y_u' = T[e_f^{\text{cut}}(\tau, T, j, 2\ell) \dots \min(n+1, j+2\ell)). \ Letting \ H = \text{root}_f(\tau, T, j) \ and \ \mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f, H}^-(\tau, T), T),$ it holds

- $$\begin{split} \bullet & |\operatorname{Pos}^{\mathrm{low}-}_{f,\ell}(j,T)| = \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l,n) \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l,n,y_u), \\ \bullet & |\operatorname{Pos}^{\mathrm{high}-}_{f,\ell}(j,T)| = \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l',n) \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l',n,y_u'). \end{split}$$

Proof. By definition, we have $j \in \text{Occ}_{\ell}(j,T)$ and $j \in \text{Occ}_{2\ell}(j,T)$. Thus, the claim follows by Lemma 8.98. \square

Query Algorithms

Proposition 8.100. Let $k \in [4..\lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $j \in \mathsf{R}^-(\tau,T)$. Given $\mathsf{CompSAPeriodic}(T)$, the value k, any position $j' \in \mathsf{Occ}_\ell(j,T)$ (resp. $j' \in \mathsf{Occ}_2(j,T)$), any $j'' \in \operatorname{Occ}_{3\tau-1}(j,T)$ satisfying $j'' = \min \operatorname{Occ}_{2\ell}(j'',T)$, and the values $\operatorname{head}_f(\tau,T,j)$, $|\operatorname{root}_f(\tau,T,j)|$, and $\exp_f^{\mathrm{cut}}(\tau,T,j,\ell) \ (\mathit{resp.}\ \exp_f^{\mathrm{cut}}(\tau,T,j,2\ell)) \ \mathit{as\ input}, \ \mathit{we\ can\ compute}\ |\mathrm{Pos}_{f,\ell}^{\mathrm{low}-}(j,T)| \ (\mathit{resp.}\ |\mathrm{Pos}_{f,\ell}^{\mathrm{high}-}(j,T)|) \ \mathit{in}$ $\mathcal{O}(\log^{2+\epsilon} n)$ time.

Proof. Let $s = \text{head}_f(\tau, T, j)$, $H = \text{root}_f(\tau, T, j)$, p = |H|, $k_1 = \exp_f^{\text{cut}}(\tau, T, j, \ell)$, and $k_2 = \exp_f^{\text{cut}}(\tau, T, j, 2\ell)$. In $\mathcal{O}(1)$ time we set $x_l := e_f^{\text{cut}}(\tau, T, j, \ell) - j = s + k_1 p$ (resp. $x_l := e_f^{\text{cut}}(\tau, T, j, 2\ell) - j = s + k_2 p$). Denote $y_u = T[e_f^{\text{cut}}(\tau, T, j, \ell) \dots \min(n+1, j+\ell))$ (resp. $y_u = T[e_f^{\text{cut}}(\tau, T, j, 2\ell) \dots \min(n+1, j+2\ell))$). Using Proposition 8.90 and the position j'' as input, in $\mathcal{O}(\log n)$ time we retrieve the pointer to the structure from Proposition 6.6 for $\operatorname{Seed}_{f,H}^-(\tau,T)$ (note that $j'' \in \mathsf{R}_{f,H}(\tau,T)$ holds by Lemma 8.50(2)), i.e., performing weighted range queries on $\mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f,H}^-(\tau,T),T)$. Note that this pointer is not null, since we assumed $j \in \mathsf{R}^-(\tau,T)$, which implies $\operatorname{Seed}_{fH}^-(\tau,T) \neq \emptyset$. Note also that using CompSANonperiodic(T), we can perform LCE_T and LCE_T queries in $\mathcal{O}(\log n)$ time, and we can access any symbol of T in $\mathcal{O}(\log n)$ time. Thus, we can compare any two substrings of T^{∞} or $\overline{T^{\infty}}$ (specified with their starting positions and lengths) in $t_{\text{cmp}} =$ $\mathcal{O}(\log n)$ time. In $\mathcal{O}(\log^{2+\epsilon} n + t_{\text{cmp}} \log n) = \mathcal{O}(\log^{2+\epsilon} n)$ time we thus compute $q := \text{weight-count}_{\mathcal{P}}(x_l, n) - t_{\text{cmp}} \log n$ weight-count_P (x_l, n, y_u) using Proposition 6.6 with $i = j' + x_l$ and $q_r = \min(n+1, j'+\ell) - i$ (resp. $q_r = \min(n+1, j'+2\ell) - i$). By Lemma 8.99, it holds $|\operatorname{Pos}_{f,\ell}^{\text{low}-}(j,T)| = q$ (resp. $|\operatorname{Pos}_{f,\ell}^{\text{high}-}(j,T)| = q$). In total, we spend $\mathcal{O}(\log^{2+\epsilon} n)$ time.

Proposition 8.101. Let $k \in [4 ... \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $j \in \mathsf{R}(\tau,T)$ be a position satisfying $j = \min \operatorname{Occ}_{2\ell}(j,T)$. Denote $s = \operatorname{head}_f(\tau,T,j)$ and $H = \operatorname{root}_f(\tau,T,j)$. Let H'be a length-s suffix of H and P be a length-2 ℓ prefix of $H'H^{\infty}$. Then, P is τ -periodic. Moreover, given $\operatorname{CompSAPeriodic}(T)$, the value k, the position j, and the values s and |H| as input, we can compute $|\operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T)|$ and $|\operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P,T)|$ in $\mathcal{O}(\log^{2+\epsilon}n)$ time.

Proof. Denote p = |H|. In $\mathcal{O}(1)$ time we compute $k_1 := \lfloor \frac{\ell - s}{p} \rfloor$, $k_2 := \lfloor \frac{2\ell - s}{p} \rfloor$, $x_l := s + k_1 p$, and $x_l' := s + k_2 p$. Denote $y_u = T[j + s ... j + s + \ell - x_l)$ and $y_u' = T[j + s ... j + s + 2\ell - x_l')$. Using Proposition 8.90 and the position j as input, in $\mathcal{O}(\log n)$ time we check if $\operatorname{Seed}_{f,H}^-(\tau,T) \neq \emptyset$, and if so, we retrieve the pointer μ_H to the structure from Proposition 6.6 for $\operatorname{Seed}_{f,H}^-(\tau,T)$, i.e., performing weighted range queries on $\mathcal{P}=$ IntStrPoints_{7 τ} (Seed_{f,H}(τ , T), T). Note that using CompSACore(T) (which is part of CompSAPeriodic(T)), we can lexicographically compare any two substrings of T^{∞} or $\overline{T^{\infty}}$ (specified with their starting positions and lengths) in $t_{\text{cmp}} = \mathcal{O}(\log n)$ time. If $\text{Seed}_{f,H}^-(\tau,T) = \emptyset$ (observe that this is possible, since we did not assume anything about $\text{type}(\tau,T,j)$), then we return $|\text{Pos}_{f,\ell}^{\text{low}^-}(P,T)| = |\text{Pos}_{f,\ell}^{\text{high}^-}(P,T)| = 0$. Otherwise, in $\mathcal{O}(\log^{2+\epsilon} n + t_{\text{cmp}} \log n) = \mathcal{O}(\log^{2+\epsilon} n)$ time we compute $q_1 := \text{weight-count}_{\mathcal{P}}(x_l,n) - \text{weight-count}_{\mathcal{P}}(x_l,n,y_u)$ and $q_2 := \text{weight-count}_{\mathcal{P}}(x_l', n) - \text{weight-count}_{\mathcal{P}}(x_l', n, y_u')$ using Proposition 6.6 first with i = j + s and $q_r = \ell - x_l$, and then with i = j + s and $q_r = 2\ell - x_l'$. By Lemma 8.97, it holds $|\operatorname{Pos}_{f,\ell}^{\text{low}-}(P, T)| = q_1$ and $|\operatorname{Pos}_{f,\ell}^{\text{high}-}(P, T)| = q_2$. In total, we spend $\mathcal{O}(\log^{2+\epsilon} n)$ time.

8.4.8 Computing the Size of Posmid

Combinatorial Properties

Lemma 8.102. Let $\ell \geq 16$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^+$ be a τ periodic pattern and let $s = \text{head}_f(\tau, P)$, $H = \text{root}_f(\tau, P)$, and p = |H|. Letting $k_{\min} = \lceil \frac{3\tau - 1 - s}{p} \rceil - 1$ and $k_{\max} = \lfloor \frac{7\tau - s}{p} \rfloor$, it holds $k_{\min} \leq \exp_f^{\text{cut}}(\tau, P, \ell) \leq \exp_f^{\text{cut}}(\tau, P, 2\ell) \leq k_{\max}$. *Proof.* Denote $k_1 = \exp_f^{\text{cut}}(\tau, P, \ell)$ and $k_2 = \exp_f^{\text{cut}}(\tau, P, 2\ell)$. We prove the three inequalities as follows.

- First, note that P being τ -periodic implies that $e(\tau,P)-1\geq 3\tau-1$, which in turn implies $\exp_f(\tau,P)=\lfloor\frac{e(\tau,P)-1-s}{p}\rfloor\geq \lfloor\frac{3\tau-1-s}{p}\rfloor\geq \lfloor\frac{3\tau-1-s}{p}\rfloor-1$. On the other hand, by $\tau=\lfloor\frac{\ell}{3}\rfloor$ it follows that $3\tau-1\leq \ell$, and hence $\lfloor\frac{\ell-s}{p}\rfloor\geq \lfloor\frac{3\tau-1-s}{p}\rfloor\geq \lfloor\frac{3\tau-1-s}{p}\rfloor-1$. Putting the two together, we thus obtain $k_{\min}=1$ and hence $\lfloor \frac{s}{p} \rfloor \geq \lfloor \frac{s}{p} \rfloor \leq \lfloor \frac{s}{p} \rfloor = 1$. I using the two together, we have $\lceil \frac{3\tau-1-s}{p} \rceil - 1 \leq \min(\exp_f(\tau,P),\lfloor \frac{\ell-s}{p} \rfloor) = k_1$.

 • Second, note that $k_1 = \min(\exp_f(\tau,P),\lfloor \frac{\ell-s}{p} \rfloor) \leq \min(\exp_f(\tau,P),\lfloor \frac{2\ell-s}{p} \rfloor) = k_2$.

 • Lastly, $k_2 = \min(\exp_f(\tau,P),\lfloor \frac{2\ell-s}{p} \rfloor) \leq \lfloor \frac{2\ell-s}{p} \rfloor \leq \lfloor \frac{7\tau-s}{p} \rfloor = k_{\max}$, where $2\ell \leq 7\tau$ follows by $\tau = \lfloor \frac{\ell}{3} \rfloor$ and
- $\ell \geq 16$.

Lemma 8.103. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^+$ be $a \ \tau$ -periodic pattern. Denote $s = \text{head}_f(\tau, P), \ H = \text{root}_f(\tau, P), \ p = |H|, \ k_1 = \exp^{\text{cut}}_{\tau}(\tau, P, \ell), \ k_2 = \exp^{\text{cut}}_{\tau}(\tau, P, \ell)$ $\exp_f^{\text{cut}}(\tau, P, 2\ell)$, and $\mathcal{I} = \text{WeightedIntervals}_{7\tau}(\text{Seed}_{f,H}^-(\tau, T), T)$. Then, it holds

$$|\operatorname{Pos}_{f,\ell}^{\operatorname{mid}}(P,T)| = \operatorname{\mathsf{mod-count}}_{\mathcal{I},p}(s,k_1,k_2).$$

Proof. Denote $k_{\min} = \lceil \frac{3\tau - 1 - s}{p} \rceil - 1$ and $k_{\max} = \lfloor \frac{7\tau - s}{p} \rfloor$. By Lemma 8.102, it holds $k_{\min} \le k_1 \le k_2 \le k_{\max}$. Thus, by applying Definition 8.70 and Lemma 8.87, we obtain $|\operatorname{Pos}_{f,\ell}^{\operatorname{mid}}(P,T)| = |\{j' \in \mathsf{R}_{f,s,H}^-(\tau,T): T\}|$ $\exp_f(\tau, T, j') \in (k_1 ... k_2]\} = \mathsf{mod\text{-}count}_{\mathcal{I}, p}(s, k_1, k_2).$

Query Algorithms

Proposition 8.104. Let $k \in [4..\lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $P \in \Sigma^+$ be a τ -periodic pattern. Given CompSAPeriodic(T), the values k, head $f(\tau, P)$, $|\operatorname{root}_f(\tau, P)|$, $\exp_f^{\mathrm{cut}}(\tau,P,\ell), \ \exp_f^{\mathrm{cut}}(\tau,P,2\ell), \ and \ some \ j \in \mathrm{Occ}(P[1\mathinner{\ldotp\ldotp} 3\tau-1],T) \ satisfying \ j = \min \mathrm{Occ}_{2\ell}(j,T), \ we \ can be considered as the constant of the constant of$ compute $|\operatorname{Pos}_{f,\ell}^{\operatorname{mid}}(P,T)|$ in $\mathcal{O}(\log^{2+\epsilon} n)$ time.

Proof. Denote $s = \text{head}_f(\tau, P)$, $H = \text{root}_f(\tau, P)$, p = |H|, $k_1 = \exp_f^{\text{cut}}(\tau, P, \ell)$, and $k_2 = \exp_f^{\text{cut}}(\tau, P, 2\ell)$. Using Proposition 8.91 and the position j as input, in $\mathcal{O}(\log n)$ time we check if $\operatorname{Seed}_{f,H}^-(\tau,T) \neq \emptyset$ (note that $j \in \mathsf{R}_{f,H}(\tau,T)$ follows by Lemma 8.50(1)), and if so, retrieve the pointer to the structure from Proposition 7.3 for Seed $_{f,H}^-(\tau,T)$, i.e., performing weighted modular constraint queries on $\mathcal{I}=$ WeightedIntervals_{7 τ}(Seed⁻_{f,H}(τ , T), T). If Seed⁻_{f,H}(τ , T) = \emptyset , then by Lemma 8.103, it holds $|\operatorname{Pos}^{\operatorname{mid}}_{f,\ell}(P,T)| = \emptyset$ $\mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k_1,k_2) = 0$, and hence we return 0. Otherwise, in $\mathcal{O}(\log^{2+\epsilon} n)$ time we compute q := $\mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k_1,k_2)$ using Proposition 7.3. By Lemma 8.103, it holds $|\mathsf{Pos}^{\mathsf{mid}}_{f,\ell}(P,T)| = q$. In total, we spend $\mathcal{O}(\log^{2+\epsilon}n)$ time.

Proposition 8.105. Let $k \in [4.. \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $j \in R(\tau,T)$. Given CompSAPeriodic(T), the values k, head_f(τ, T, j), $|\text{root}_f(\tau, T, j)|$, $\exp_f^{\text{cut}}(\tau, T, j, \ell)$, $\exp_f^{\text{cut}}(\tau, T, j, 2\ell)$, and some $j' \in \text{Occ}_{3\tau-1}(j,T)$ satisfying $j' = \min \text{Occ}_{2\ell}(j',T)$ as input, we can compute $|\text{Pos}_{f,\ell}^{\text{mid}-}(j,T)|$ in $\mathcal{O}(\log^{2+\epsilon} n)$ time.

Proof. Denote $s = \text{head}_f(\tau, T, j)$, $H = \text{root}_f(\tau, T, j)$, p = |H|, $k_1 = \exp_f^{\text{cut}}(\tau, T, j, \ell)$, $k_2 = \exp_f^{\text{cut}}(\tau, T, j, 2\ell)$, and P = T[j ... n]. By $j \in R(\tau, T)$, P is τ -periodic. We then also have $j' \in Occ_{3\tau-1}(j, T) = Occ(P[1...3\tau -$ 1], T) and head $f(\tau, P) = s$, root $f(\tau, P) = H$, $\exp_f^{\text{cut}}(\tau, P, \ell) = k_1$, and $\exp_f^{\text{cut}}(\tau, P, 2\ell) = k_2$. Thus, using s, p, k_1 , k_2 , and j' as input to Proposition 8.104, we can compute $|\operatorname{Pos}_{f,\ell}^{\text{mid}}(j, T)| = |\operatorname{Pos}_{f,\ell}^{\text{mid}}(T[j \dots n], T)|$ $|\operatorname{Pos}_{f,\ell}^{\operatorname{mid}^{-}}(P,T)|$ in $\mathcal{O}(\log^{2+\epsilon} n)$ time.

Proposition 8.106. Let $k \in [4 ... \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $j \in \mathsf{R}(\tau,T)$ be a position satisfying $j = \min \operatorname{Occ}_{2\ell}(j,T)$. Denote $s = \operatorname{head}_f(\tau,T,j)$ and $H = \operatorname{root}_f(\tau,T,j)$. Let H'be a length-s suffix of H and P be a length-2 ℓ prefix of $H'H^{\infty}$. Then, P is τ -periodic. Moreover, given CompSAPeriodic(T), the value k, the position j, and the values head $f(\tau, T, j)$ and $|\operatorname{root}_f(\tau, T, j)|$ as input, we compute $|\operatorname{Pos}_{f,\ell}^{\operatorname{mid}^-}(P,T)|$ in $\mathcal{O}(\log^{2+\epsilon} n)$ time.

Proof. Denote p = |H|. In $\mathcal{O}(1)$ time, we compute $k_1 = \lfloor \frac{\ell - s}{p} \rfloor$ and $k_2 = \lfloor \frac{2\ell - s}{p} \rfloor$. By Lemma 8.57, P is τ -periodic and it holds head $f(\tau, P) = s$, $|\operatorname{root}_f(\tau, P)| = p$, $\exp_f^{\operatorname{cut}}(\tau, P, \ell) = k_1$, and $\exp_f^{\operatorname{cut}}(\tau, P, 2\ell) = k_2$. It also implies that $j \in \text{Occ}_{3\tau-1}(P,T)$. Thus, using j, s, p, k_1 , and k_2 as input to Proposition 8.104, we compute $|\operatorname{Pos}_{f,\ell}^{\operatorname{mid}}(\tau,P)|$ in $\mathcal{O}(\log^{2+\epsilon} n)$ time.

8.4.9 Computing the Exponent

Combinatorial Properties

Lemma 8.107. Let $\ell \in [16 ...n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^+$ be a τ -periodic pattern. Then, it holds $\operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T) = \{\operatorname{SA}[i]: i \in (b ...e]\}$ (resp. $\operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P,T) = \{\operatorname{SA}[i]: i \in (b ...e]\}$), where $b = \operatorname{RangeBeg}_{\ell}(P,T)$ (resp. $b = \operatorname{RangeBeg}_{2\ell}(P,T)$) and $e = b + \delta_{f,\ell}^{\operatorname{low}^-}(P,T)$ (resp. $e = b + \delta_{f,\ell}^{\operatorname{high}^-}(P,T)$).

Proof. We prove the claim only for $\operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T)$. The proof for $\operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P,T)$ is nearly identical, except we replace $\operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T)$, $\delta_{f,\ell}^{\operatorname{low}^-}(P,T)$, ℓ , k_1 , and $\operatorname{exp}_f^{\operatorname{cut}}(\tau,P,\ell)$ with $\operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P,T)$, $\delta_{f,\ell}^{\operatorname{high}^-}(P,T)$, 2ℓ , k_2 , and $\operatorname{exp}_f^{\operatorname{cut}}(\tau,P,2\ell)$, respectively.

Denote m = |P| and $P' = P[1 \dots \min(m, \ell)]$. First, we prove that $\operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T) \subseteq \{\operatorname{SA}[i] : i \in (b \dots n]\}$. Consider any $j' \in \operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T)$ and let $i' \in [1 \dots n]$ be such that $\operatorname{SA}[i'] = j'$. By definition of $\operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T)$, it holds $T[j' \dots n] \succeq P$ or $\operatorname{lcp}(P,T[j' \dots n]) \geq \ell$. By Lemma 8.8(1), this implies that $T[j' \dots n] \succeq P'$. On the other hand, by Lemma 8.8(2), for every $j \in [1 \dots n]$, $T[j \dots n] \prec P$ and $\operatorname{lcp}(P,T[j \dots n]) < \ell$ holds it and only if $T[j \dots n] \prec P'$. Thus, applying the definition of b, we obtain $b = |\{j \in [1 \dots n] : T[j \dots n] \prec P$ and $\operatorname{lcp}(P,T[j \dots n]) < \ell\}| = |\{j \in [1 \dots n] : T[j \dots n] \prec P'\}|$. By definition of suffix array, position in the last set must occupy its initial segment. We thus have $\{\operatorname{SA}[i] : i \in [1 \dots b]\} = \{j \in [1 \dots n] : T[j \dots n] \prec P'\}$. Consequently, $j' \notin \{\operatorname{SA}[i] : i \in [1 \dots b]\}$, i.e., $i' \in (b \dots n]$. We have thus proved $\operatorname{Pos}_{f,\ell}^{\operatorname{low}}(P,T) \subseteq \{\operatorname{SA}[i] : i \in (b \dots n]\}$.

Next, we show that for every $i' \in (b+1..n]$, $SA[i'] \in Pos_{f,\ell}^{low-}(P,T)$ implies $SA[i'-1] \in Pos_{f,\ell}^{low-}(P,T)$. Denote $s = head_f(\tau, P)$, $H = root_f(\tau, P)$, and p = |H|.

- As noted above i'-1>b implies that $T[SA[i'-1]..n]\succeq P'$. By Lemma 8.8(1), this implies $T[SA[i'-1]..n]\succeq P$ or $lcp(P,T[SA[i'-1]..n])\geq \ell$.
- First, note that $SA[i'] \in Pos_{f,\ell}^{low}(P,T)$ implies that $SA[i'] \in R_{f,s,H}^-(\tau,T)$. By Lemma 8.50(1), we thus have $lcp(P,T[SA[i']..n]) \geq 3\tau 1$. Denote h = lcp(T[SA[i'-1]..n],T[SA[i']..n]). We show that $h \geq 3\tau 1$. Assume the opposite and consider two cases:
 - First, assume $n \mathrm{SA}[i'-1] + 1 = h$. Then, $T[\mathrm{SA}[i'-1] \dots n]$ is a proper prefix of $T[\mathrm{SA}[i'] \dots \mathrm{SA}[i'] + 3\tau 1) = P[1 \dots 3\tau 1]$. Note, however, that $|P'| = \min(m,\ell) \ge 3\tau 1$, since $\ell \ge 3\tau 1$ follows by the definition, and $m \ge 3\tau 1$ follows since we assumed that P is τ -periodic (Definition 8.1). Since P' is a prefix of P, we thus obtain that $T[\mathrm{SA}[i'-1] \dots n]$ is a proper prefix of P'. Thus, $T[\mathrm{SA}[i'-1] \dots n] \prec P'$. However, as observed above, this implies $i'-1 \in [1 \dots b]$, contradicting the assumption $i' \in (b+1 \dots n]$.
 - Let us now assume $n \mathrm{SA}[i'-1] + 1 > h$. Then, $\mathrm{SA}[i'-1]$ being in the suffix array earlier than $\mathrm{SA}[i']$ implies that $n \mathrm{SA}[i'] + 1 > h$ and $T[\mathrm{SA}[i'-1] + h] \prec T[\mathrm{SA}[i'] + h]$. Note, however, since by $h < 3\tau 1$, we have $T[\mathrm{SA}[i'] \ldots \mathrm{SA}[i'] + h] = P'[1 \ldots h + 1]$. We therefore again obtain $T[\mathrm{SA}[i'-1] \ldots n] \prec P'$, which by the above implies $i'-1 \in [1 \ldots b]$, contradicting the assumption $i' \in (b+1 \ldots n]$.

We thus obtain $h \geq 3\tau - 1$. By Lemma 8.50(2), this implies head_f $(\tau, T, SA[i'-1]) = \text{head}_f(\tau, T, SA[i']) = s$ and $\text{root}_f(\tau, T, SA[i'-1]) = \text{root}_f(\tau, T, SA[i']) = H$. Lastly, observe that we must also have $\text{type}(\tau, T, SA[i'-1]) = -1$, since otherwise by Lemma 8.52(2), we would have $T[SA[i'-1] : n] \succ T[SA[i'] : n]$, contradicting the definition of suffix array. We have thus proved $SA[i'-1] \in \mathbb{R}^-_{f,s,H}(\tau, T)$.

• Denote $k_1 = \exp_f^{\operatorname{cut}}(\tau, P, \ell)$. We show that $\exp_f(\tau, T, \operatorname{SA}[i'-1]) = k_1$. To this end, we first prove that P' is τ -periodic and it holds $\operatorname{head}_f(\tau, P') = s$, $\operatorname{root}_f(\tau, P') = H$, and $\exp_f(\tau, P') = k_1$. As noted above, it holds $|P'| \geq 3\tau - 1$. Moreover, P' is a prefix of P. By Lemma 8.45, we thus obtain that P' is τ -periodic and it holds $\operatorname{head}_f(\tau, P') = \operatorname{head}_f(\tau, P) = s$ and $\operatorname{root}_f(\tau, P') = \operatorname{root}_f(\tau, P) = H$. Lastly, by Lemma 8.47, we have $k_1 = \exp_f^{\operatorname{cut}}(\tau, P, \ell) = \exp_f(\tau, P[1 \dots \min(m, \ell)]) = \exp_f(\tau, P')$. We have thus proved that P' is τ -periodic and it holds $\operatorname{head}_f(\tau, P') = s$, $\operatorname{root}_f(\tau, P') = H$, and $\exp_f(\tau, P') = k_1$. Observe now that by Lemma 8.52(3), we have $e(\tau, T, \operatorname{SA}[i'-1]) - \operatorname{SA}[i'-1] \leq e(\tau, T, \operatorname{SA}[i']) - \operatorname{SA}[i']$. Thus, $\exp_f(\tau, T, \operatorname{SA}[i'-1]) = \lfloor \frac{e(\tau, T, \operatorname{SA}[i'-1]) - \operatorname{SA}[i'-1] - s}{p} \rfloor \leq \lfloor \frac{e(\tau, T, \operatorname{SA}[i']) - s}{p} \rfloor = \exp_f(\tau, T, \operatorname{SA}[i']) = k_1$ (the last equality follows by $\operatorname{SA}[i'] \in \operatorname{Pos}_{f,\ell}^{low}(P, T)$). Consequently, to show $\exp_f(\tau, T, \operatorname{SA}[i'-1]) = k_1$,

it remains to prove that we cannot have $\exp_f(\tau,T,\operatorname{SA}[i'-1]) < k_1$. Let us thus assume that this holds. Then, $e(\tau,T,\operatorname{SA}[i'-1]) - \operatorname{SA}[i'-1] = \operatorname{head}_f(\tau,T,\operatorname{SA}[i'-1]) + \exp_f(\tau,T,\operatorname{SA}[i'-1])|\operatorname{root}_f(\tau,T,\operatorname{SA}[i'-1])| + \operatorname{tail}_f(\tau,T,\operatorname{SA}[i'-1]) \le s + (k_1-1)p + \operatorname{tail}_f(\tau,T,\operatorname{SA}[i'-1]) < s + k_1p \le s + k_1p + \operatorname{tail}_f(\tau,P') = \operatorname{head}_f(\tau,P') + \exp_f(\tau,P')|\operatorname{root}_f(\tau,P')| + \operatorname{tail}_f(\tau,P') = e(\tau,P') - 1$. By Lemma 8.51(3), we then have $T[\operatorname{SA}[i'-1] \dots n] \prec P'$. As noted above, this implies $i'-1 \in [1 \dots b]$, contradicting $i' \in (b+1 \dots n]$. We have thus proved that $\exp_f(\tau,T,\operatorname{SA}[i'-1]) = k_1$.

Combining the above conditions implies $\mathrm{SA}[i'-1] \in \mathrm{Pos}^{\mathrm{low}-}_{f,\ell}(P,T)$. Consequently, $\mathrm{Pos}^{\mathrm{low}-}_{f,\ell}(P,T) = \{\mathrm{SA}[i]: i \in (b\mathinner{.\,.} b+|\mathrm{Pos}^{\mathrm{low}-}_{f,\ell}(P,T)|]\} = \{\mathrm{SA}[i]: i \in (b\mathinner{.\,.} e]\}$. \square

Corollary 8.108. Let $\ell \in [16 \dots n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^+$ be a τ -periodic pattern. For any $i \in [1 \dots n]$, $\mathrm{SA}[i] \in \mathrm{Pos}^{\mathrm{low}-}_{f,\ell}(P,T)$ (resp. $\mathrm{SA}[i] \in \mathrm{Pos}^{\mathrm{high}-}_{f,\ell}(P,T)$) holds if and only if $\mathrm{RangeBeg}_{\ell}(P,T) < i \leq \mathrm{RangeBeg}_{\ell}(P,T) + \delta^{\mathrm{low}-}_{f,\ell}(P,T)$ (resp. $\mathrm{RangeBeg}_{2\ell}(P,T) < i \leq \mathrm{RangeBeg}_{2\ell}(P,T) + \delta^{\mathrm{high}-}_{f,\ell}(P,T)$).

Lemma 8.109. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^+$ be a τ -periodic pattern satisfying type $(\tau, P) = -1$. Denote $s = \text{head}_f(\tau, P)$, $H = \text{root}_f(\tau, P)$, $k_1 = \exp_f^{\text{cut}}(\tau, P, \ell)$, $b = \text{RangeBeg}_\ell(P, T)$, and $\delta = \delta_{f,\ell}^{\text{low}}(P, T)$. For any $k \geq k_1$, let $E_k := \bigcup_{t \in (k_1..k]} \mathsf{R}_{f,s,t,H}^-(\tau, T)$ and $m_k = |E_k|$. It holds $E_k = \{ \mathsf{SA}[i] : i \in (b + \delta ...b + \delta + m_k] \}$.

Proof. Denote p = |H|, $i' = b + \delta$, and $B = \{j \in [1 ... n] : T[j ... n] \prec P \text{ and } lcp(P, T[j ... n]) < \ell\}$. In the proof of Lemma 8.107, we observed that $B = \{SA[i] : i \in [1 ... b]\}$. We proceed in two steps.

First, we show that $E_k \subseteq \{SA[i'']: i'' \in (i' ... n]\}$. The claim holds trivially for $k = k_1$ since $E_{k_1} = \emptyset$. Let us thus assume $k > k_1$. Let $j \in E_k$ and let $i \in [1...n]$ be such that SA[i] = j. Note that by definition, we have $\exp_f(\tau, P) \ge k_1$. Consider two cases:

- First, assume that $\exp_f(\tau,P) = k_1$. Then, $e(\tau,P) 1 = \operatorname{head}_f(\tau,P) + \exp_f(\tau,P) | \operatorname{root}_f(\tau,P)| + \operatorname{tail}_f(\tau,P) = s + k_1p + \operatorname{tail}_f(\tau,P) < s + (k_1+1)p \le s + (k_1+1)p + \operatorname{tail}_f(\tau,T,j) \le \operatorname{head}_f(\tau,T,j) + \exp_f(\tau,T,j) | \operatorname{root}_f(\tau,T,j)| + \operatorname{tail}_f(\tau,T,j) = e(\tau,T,j) j$. By Lemma 8.51(1) and Lemma 8.51(5), we thus have $P \prec T[j \ldots n]$. Consequently, $j \notin B$, and hence $i \in (b \ldots n]$. Note, however, that since for every $j' \in \operatorname{Pos}_{f,\ell}^{\text{low}-}(P,T)$ we have $\exp_f(\tau,T,j') = k_1 < \exp_f(\tau,T,j)$, it also holds $j \notin \operatorname{Pos}_{f,\ell}^{\text{low}-}(P,T)$. By Corollary 8.108, we thus have $i \notin (b \ldots b + \delta] = (b \ldots i']$. Consequently, $i \in (i' \ldots n]$. Thus, $j \in \{\operatorname{SA}[i''] : i'' \in (i' \ldots n]\}$.
- Let us now assume $\exp_f(\tau, P) > k_1$. We begin by observing that the assumption $\exp_f(\tau, P) > k_1$ implies that, letting H' be a length-s suffix of H, the string $H'H^{k_1+1}$ is a prefix of P. On the other hand, by definition of E_k , we have $\exp_f(\tau, T, j) > k_1$. Thus, $H'H^{k_1+1}$ is also a prefix of $T[j \dots n]$, and hence $\operatorname{lcp}(P, T[j \dots n]) \ge s + (k_1 + 1)p$. Observe now that the assumption $k_1 = \min(\exp_f(\tau, P), \lfloor \frac{\ell s}{P} \rfloor) < \exp_f(\tau, P)$ implies that $k_1 = \lfloor \frac{\ell s}{p} \rfloor$. Consequently, $s + (k_1 + 1)p = s + (\lfloor \frac{\ell s}{p} \rfloor + 1)p \ge s + (\frac{\ell s}{p})p = \ell$. Thus, $\operatorname{lcp}(P, T[j \dots n]) \ge \ell$, and hence $j \notin B$. Similarly, as in the first case, we also have $j \notin \operatorname{Pos}_{f,\ell}^{\text{low}}(P, T)$. Thus, again $i \in (i' \dots n]$ and hence $j \in \{\operatorname{SA}[i''] : i'' \in (i' \dots n]\}$.

Second, we prove that for every $i'' \in (i'+1 \dots n]$, $\mathrm{SA}[i''] \in E_k$ implies $\mathrm{SA}[i''-1] \in E_k$. We begin by proving that $\mathrm{SA}[i''-1] \in \mathsf{R}_{f,s,H}(\tau,T)$ and $\exp_f(\tau,T,\mathrm{SA}[i''-1]) \geq k_1$. Observe that $i''-1>i'\geq b$ implies that $\mathrm{SA}[i''-1] \not\in B$, i.e., that it holds $T[\mathrm{SA}[i''-1] \dots n] \succeq P$ or $\mathrm{lcp}(P,T[\mathrm{SA}[i''-1] \dots n]) \geq \ell$. Consider two cases:

- First, assume $lcp(P, T[SA[i''-1]..n]) \ge \ell$. By $\ell \ge 3\tau 1$ and Lemma 8.50(1), we then obtain $SA[i''-1] \in R_{f,s,H}(\tau,T)$. Denote $P' = P[1..\ell]$. Note that by Lemma 8.45, P' is τ -periodic and we have $head_f(\tau,P') = s$ and $root_f(\tau,P') = H$. Observe that $e(\tau,T,SA[i''-1]) SA[i''-1] = p + lcp(T[SA[i''-1]..n], T[SA[i''-1]+p..n]) \ge p + lcp(T[SA[i''-1]..SA[i''-1]+\ell), T[SA[i''-1]+p..n]) = p + lcp(P'[1..\ell], P'[1+p..\ell]) = e(\tau,P')-1$. Consequently, $exp_f(\tau,T,SA[i''-1]) = \lfloor \frac{e(\tau,T,SA[i''-1])-SA[i''-1]-s}{p} \rfloor \ge \lfloor \frac{e(\tau,P')-1-s}{p} \rfloor = exp_f(\tau,P')$. It remains to note that by Lemma 8.47, $exp_f(\tau,P') = exp_f^{cut}(\tau,P,\ell) = k_1$. Hence, $exp_f(\tau,T,SA[i''-1]) \ge k_1$.
- $\exp_f(\tau, P') = \exp_f^{\text{cut}}(\tau, P, \ell) = k_1. \text{ Hence, } \exp_f(\tau, T, \text{SA}[i''-1]) \ge k_1.$ Let us now assume $T[\text{SA}[i''-1] ...n] \succeq P$. If T[SA[i''-1] ...n] = P, then by Lemma 8.50(1), we immediately obtain that $\text{SA}[i''-1] \in \mathsf{R}_{f,s,H}(\tau,T)$ and also by definition $\exp_f(\tau, T, \text{SA}[i''-1]) = \mathsf{R}_{f,s,H}(\tau,T)$

 $\exp_f(\tau, T[SA[i''-1]..n]) = \exp_f(\tau, P) \ge k_1$. Let us thus assume T[SA[i''-1]..n] > P. Denote X = P, Y = T[SA[i''-1]..n], and Z = T[SA[i'']..n]. We then have $X \prec Y \prec Z$. Note now that both X and Z are τ -periodic and it holds head $f(\tau, X) = \text{head}_f(\tau, Z)$ and root $f(\tau, X) = \text{root}_f(\tau, Z)$. Thus, by Lemma 8.45, we have $lcp(X,Z) \ge 3\tau - 1$. Observe now that since lcp(X,Z) = min(lcp(X,Y), lcp(Y,Z))holds for any strings satisfying $X \prec Y \prec Z$, we thus must have that lcp(X,Y) = lcp(P,T[SA[i''- $1]...n] \ge 3\tau - 1$. Thus, by Lemma 8.50(1), we have $SA[i''-1] \in R_{f,s,H}(\tau,T)$. Let H' be a lengths suffix of H. By $\exp_f(\tau, P) \ge \exp_f^{\text{cut}}(\tau, P, \ell) = k_1$, $H'H^{k_1}$ is a prefix of P. On the other hand, $SA[i''] \in E_k$ implies $\exp_f(\tau, T, SA[i'']) > k_1$. Thus, $H'H^{k_1}$ is also a prefix of T[SA[i''] ... n]. Consequently, $lcp(X,Z) \ge s + k_1p$. By the same argument as above, we thus also have lcp(X,Y) = lcp(P,T[SA[i''-1]) $[1] \dots n] \ge s + k_1 p$, i.e., $H'H^{k_1}$ is also a prefix of $T[SA[i''-1] \dots n]$. Since $SA[i''-1] \in R_{f,s,H}(\tau,T)$, this implies $e(\tau, T, \mathrm{SA}[i''-1]) - \mathrm{SA}[i''-1] = p + \mathrm{lcp}(T[\mathrm{SA}[i''-1] \dots n], T[\mathrm{SA}[i''-1] + p \dots n]) \ge s + k_1 p$, and consequently, $\exp_f(\tau, T, \mathrm{SA}[i''-1]) = \lfloor \frac{e(\tau, T, \mathrm{SA}[i''-1]) - \mathrm{SA}[i''-1] - s}{p} \rfloor \ge \lfloor \frac{k_1 p}{p} \rfloor = k_1$.

We have thus proved that $SA[i''-1] \in R_{f,s,H}(\tau,T)$ and $\exp_f(\tau,T,SA[i''-1]) \ge k_1$. Next, observe that we must have $\operatorname{type}(\tau, T, \operatorname{SA}[i''-1]) = -1$, since otherwise by $\operatorname{type}(\tau, T, \operatorname{SA}[i'']) = -1$ and Lemma 8.52(2), we would have $T[SA[i''-1]..n] \succ T[SA[i'']..n]$, contradicting the definition of the suffix array. We have thus proved that $SA[i''-1] \in R_{f,s,H}^{-}(\tau,T)$. Recall now the definition of $Pos_{f,\ell}^{low}(P,T)$. Observe that almost all conditions are satisfied for SA[i''-1], except possibly the condition on the exponent. Since, however, we assumed i''-1>i', by Lemma 8.107 we have $SA[i''-1] \notin Pos_{f,\ell}^{low}(P,T)$. Consequently, we must have $\exp_f(\tau, T, \mathrm{SA}[i''-1]) \neq k_1$, and hence $\exp_f(\tau, T, \mathrm{SA}[i''-1]) > k_1$. Lastly, we prove that $\exp_f(\tau, T, \operatorname{SA}[i''-1]) \le k$. Suppose $\exp_f(\tau, T, \operatorname{SA}[i''-1]) \ge k+1$. Then, it holds $e(\tau, T, \operatorname{SA}[i''-1]) - e(\tau, T, \operatorname{SA}[i''-1]) = k+1$. $\mathrm{SA}[i''-1] = \mathrm{head}_f(\tau, T, \mathrm{SA}[i''-1]) + \exp_f(\tau, T, \mathrm{SA}[i''-1]) | \mathrm{root}_f(\tau, T, \mathrm{SA}[i''-1]) | + \mathrm{tail}_f(\tau, T, \mathrm{SA}[i''-1]) \geq \mathrm{tail}_f(\tau, T, \mathrm{SA}[i''-1]) | + \mathrm$ $s + (k+1)p > s + kp + \operatorname{tail}_f(\tau, T, \operatorname{SA}[i'']) \ge s + \exp_f(\tau, T, \operatorname{SA}[i''])p + \operatorname{tail}_f(\tau, T, \operatorname{SA}[i'']) = \operatorname{head}_f(\tau, T, \operatorname{SA}[i'']) + \operatorname{head}_f(\tau, T, \operatorname{SA}[i'']) = \operatorname{head}_f(\tau, T, \operatorname{SA}[i'']) + \operatorname{head}_f(\tau, T, \operatorname{SA}[i'']) = \operatorname{head}_f(\tau, T, \operatorname{SA}[i'']) = \operatorname{head}_f(\tau, T, \operatorname{SA}[i'']) + \operatorname{head}_f(\tau, T, \operatorname{SA}[i'']) = \operatorname{head}_f(\tau,$ $\exp_f(\tau, T, \operatorname{SA}[i'']) |\operatorname{root}_f(\tau, T, \operatorname{SA}[i''])| + \operatorname{tail}_f(\tau, T, \operatorname{SA}[i'']) = e(\tau, T, \operatorname{SA}[i'']) - \operatorname{SA}[i'']. \text{ By SA}[i'' - 1], \operatorname{SA}[i''] \in \operatorname{SA}[i'']$ $R_{f,s,H}^{-}(\tau,T)$ and Lemma 8.52(3) this, however, implies $T[SA[i''-1]..n] \succ T[SA[i'']..n]$, contradicting the definition of the suffix array. Thus, it holds $\exp_f(\tau, T, SA[i''-1]) \leq k$. We have therefore proved $\exp_f(\tau, T, \operatorname{SA}[i''-1]) \in (k_1 \dots k]$, which concludes the proof of $\operatorname{SA}[i''-1] \in E_k$.

Putting the above steps together implies that $E_k = \{SA[i''] : i'' \in (i' ... i' + m_k]\}.$

Lemma 8.110. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^+$ be a τ periodic pattern such that it holds $\operatorname{type}(\tau, P) = -1$, $e(\tau, P) \leq |P|$, and $\operatorname{exp}_f^{\operatorname{cut}}(\tau, P, \ell) < \operatorname{exp}_f(\tau, P) \leq \lfloor \frac{7\tau - s}{n} \rfloor$, where $s = \operatorname{head}_f(\tau, P)$, $H = \operatorname{root}_f(\tau, P)$, and p = |H|. Denote $\mathcal{I} = \operatorname{WeightedIntervals}_{7\tau}(\operatorname{Seed}_{f,H}^-(\tau, T), T)$, $k_1 = \exp_f^{\operatorname{cut}}(\tau, P, \ell)$, $c = \operatorname{mod-count}_{\mathcal{I},p}(s, k_1)$, and $i' = \operatorname{RangeBeg}_\ell(P, T) + \delta_{f,\ell}^{\operatorname{low}}^-(P, T)$. Let $i \in [1 \dots n]$ be such that $SA[i] \in Occ(P,T)$. Then, it holds:

- $\begin{array}{l} \bullet \ \ c+(i-i') \in [1 \mathinner{.\,.} \bmod \text{-} \mathtt{count}_{\mathcal{I},p}(s)], \\ \bullet \ \ \exp_f(\tau,P) = \bmod \text{-} \mathtt{select}_{\mathcal{I},p}(s,c+(i-i')). \end{array}$

Proof. Denote $k = \exp_f(\tau, P)$. The proof consists of two parts.

In the first step, we prove that for every $j' \in \text{Occ}(P,T)$, it holds $j' \in \mathsf{R}^-_{f,s,k,H}(\tau,T)$. Recall that $e(\tau,P) \leq \mathsf{R}^-_{f,s,k,H}(\tau,T)$. |P|, or equivalently, $|P| > e(\tau, P) - 1$. Consequently, $j' \in \text{Occ}(P, T)$ implies $\text{lcp}(P, T[j' ... n]) = |P| > e(\tau, P) - 1$ 1. From Lemma 8.50(1) we therefore obtain that $head_f(\tau, T, j') = head_f(\tau, P)$, $root_f(\tau, T, j') = root_f(\tau, P)$, $\operatorname{type}(\tau, T, j') = \operatorname{type}(\tau, P) = -1$, and $\exp_f(\tau, T, j') = \exp_f(\tau, P) = k$. In other words, $j' \in \mathsf{R}^-_{f.s.k.H}(\tau, T)$.

In the second step, we prove that it holds $c + (i - i') \in [1 \dots \mathsf{mod\text{-}count}_{\mathcal{I},p}(s)]$ (i.e., that $\mathsf{mod\text{-}select}_{\mathcal{I},p}(s,c+1)$) (i-i') is well-defined) and $k = \mathsf{mod}\text{-select}_{\mathcal{I},p}(s,c+(i-i'))$ (i.e., the main claim). Recall that $k_1 < k \leq \lfloor \frac{7\tau - s}{n} \rfloor$. For any $k' \geq k_1$, let $E_{k'} = \bigcup_{t \in (k_1...k']} \mathsf{R}^-_{f,s,t,H}(\tau,T)$ and $m_{k'} = |E_{k'}|$. Note that by $k_1 < k$, the sets E_k and E_{k-1} are well-defined and it holds $\mathsf{R}^-_{f,s,k,H}(\tau,T) = E_k \setminus E_{k-1}$. Thus, by the first step, we have $\mathsf{SA}[i] \in E_k \setminus E_{k-1}$. On the other hand, by Lemma 8.109, it holds $E_{k-1} = \{\mathsf{SA}[i''] : i'' \in (i' ...i' + m_{k-1}]\}$ and $E_k = \{\mathsf{SA}[i''] : i'' \in (i' ...i' + m_{k-1}]\}$ $i'' \in (i' \dots i' + m_k]$. Thus, we obtain $i \in (i' + m_{k-1} \dots i' + m_k]$. Observe now that by Lemmas 8.87 and 8.102 we have $m_{k-1} = |E_{k-1}| = |\{j' \in \mathsf{R}^-_{f,s,H}(\tau,T) : \exp_f(\tau,T,j') \in (k_1 \dots k-1]\}| = \mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k_1,k-1) = |E_{k-1}| = |E_{k-1}|$ 1). By Lemma 7.2(1), we thus obtain $m_{k-1} = \text{mod-count}_{\mathcal{I},p}(s,k_1,k-1) = \text{mod-count}_{\mathcal{I},p}(s,k-1)$ $\mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k_1) = \mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k-1) - c.$ Analogously, $m_k = \mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k) - c.$ We thus obtain $i \in (i' + \mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k-1) - c ... i' + \mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k) - c]$, or equivalently, $c + (i-i') \in$

 $(\mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k-1)\dots\mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k)]$. This implies $c+(i-i')\in[1\dots\mathsf{mod\text{-}count}_{\mathcal{I},p}(s)]$ and, by definition of the weighted modular constraint selection queries (see Section 7), $k = \mathsf{mod}\text{-select}_{\mathcal{I},p}(s,c+(i-i'))$.

Lemma 8.111. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $i \in [1..n]$ be such that $SA[i] \in R^-(\tau, T)$ and $e(\tau, T, SA[i]) - SA[i] < 2\ell$. Denote $b = RangeBeg_{\ell}(SA[i], T), \delta = 0$ $\delta_{f,\ell}^{\text{low}-}(\text{SA}[i],T)$, and $k_1 = \exp_f^{\text{cut}}(\tau,T,\text{SA}[i],\ell)$. If $i \leq b + \delta$, then $\exp_f(\tau,T,\text{SA}[i]) = k_1$. Otherwise, letting $H = \operatorname{root}_f(\tau, T, \operatorname{SA}[i]), \ p = |H|, \ s = \operatorname{head}_f(\tau, T, \operatorname{SA}[i]), \ \mathcal{I} = \operatorname{WeightedIntervals}_{\tau\tau}(\operatorname{Seed}_{f,H}^-(\tau, T), T), \ and$ $c = \mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k_1), it \ holds:$

- $c + (i (b + \delta)) \in [1 \dots \mathsf{mod\text{-}count}_{\mathcal{I},p}(s)],$
- $\exp_f(\tau, T, SA[i]) = \mathsf{mod}\text{-select}_{\mathcal{I}, p}(s, c + (i (b + \delta))).$

Proof. Denote P = T[SA[i] ... n]. First, observe that since we assumed $SA[i] \in R^-(\tau, T)$, it holds that P is τ -periodic, type $(\tau, P) = -1$, and $e(\tau, P) \leq |P|$. By Lemma 8.50(1), we also have head $f(\tau, P) = s$ and $\operatorname{root}_f(\tau, P) = H$. Next, note that by definition it holds:

- RangeBeg_{ℓ}(SA[i], T) = RangeBeg_{ℓ}(T[SA[i]..n], T) = RangeBeg_{ℓ}(P, T),
- $\delta_{f,\ell}^{\text{low}-}(\text{SA}[i],T) = \delta_{f,\ell}^{\text{low}-}(T[\text{SA}[i],..n],T) = \delta_{f,\ell}^{\text{low}-}(P,T)$, and $\exp_f^{\text{cut}}(\tau,T,\text{SA}[i],\ell) = \exp_f^{\text{cut}}(\tau,T[\text{SA}[i]..n],\ell) = \exp_f^{\text{cut}}(\tau,P,\ell)$.

Consequently, we have RangeBeg_{ℓ}(P,T) = b, $\delta_{f,\ell}^{low-}(P,T) = \delta$, and $\exp_f^{cut}(\tau,P,\ell) = k_1$. By Lemma 8.107, we thus have $\operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T) = \{\operatorname{SA}[t] : t \in (b ... b + \delta]\}.$

Let us first assume that it holds $i \le b + \delta$. By definition, it holds $SA[i] \in Occ_{\ell}(SA[i], T)$. Thus, b < i. On the other hand, we have $i \leq b + \delta$ by the assumption. Thus, by the above, it holds $SA[i] \in Pos_{f,\ell}^{low}(P,T)$. By definition of $\operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T)$, this implies $\exp_f(\tau,T,\operatorname{SA}[i]) = \exp_f^{\operatorname{cut}}(\tau,P,\ell) = k_1$.

Let us now assume $i > b + \delta$. By the above characterization of $\operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T)$, this implies $\operatorname{SA}[i] \not\in \operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T)$. We now show $k_1 < \exp_f(\tau,P)$. By definition, $\exp_f^{\operatorname{cut}}(\tau,P,\ell) = \min(\exp_f(\tau,P), \lfloor \frac{\ell-s}{p} \rfloor) \leq \operatorname{Pos}_{f,\ell}^{\operatorname{low}^-}(P,T)$. $\exp_f(\tau, P)$. Thus, $k_1 \leq \exp_f(\tau, P)$. Suppose that it holds $k_1 = \exp_f(\tau, P)$. Then, $SA[i] \in R_{f,s,H}^{-1}(\tau, T)$ and $\exp_f(\tau, T, SA[i]) = \exp_f(\tau, T[SA[i] ... n]) = \exp_f(\tau, P) = k_1 = \exp_f^{\text{cut}}(\tau, P, \ell)$. Moreover, we then have and $\exp_f(\tau, T, \operatorname{SA}[i]) = \exp_f(\tau, T[\operatorname{SA}[i] \dots n]) = \exp_f(\tau, P) = k_1 = \exp_f(\tau, P, \ell)$. Moreover, we then have $T[\operatorname{SA}[i] \dots n] = P$. Thus, by definition of $\operatorname{Pos}_{f,\ell}^{\text{low}-}(P,T)$, we have $\operatorname{SA}[i] \in \operatorname{Pos}_{f,\ell}^{\text{low}-}(P,T)$, which contradicts our earlier observation. Consequently, we have $k_1 < \exp_f(\tau, P)$. On the other hand, observe that $e(\tau, T, \operatorname{SA}[i]) - \operatorname{SA}[i] < 2\ell$ implies $\exp_f(\tau, P) = \exp_f(\tau, T[\operatorname{SA}[i] \dots n]) = \exp_f(\tau, T, \operatorname{SA}[i]) = \lfloor \frac{e(\tau, T, \operatorname{SA}[i]) - \operatorname{SA}[i] - s}{p} \rfloor \leq \lfloor \frac{2\ell - s}{p} \rfloor \leq \lfloor \frac{\tau - s}{p} \rfloor$, where $2\ell \leq 7\tau$ follows by $\tau = \lfloor \frac{\ell}{3} \rfloor$ and $\ell \geq 16$. Thus, it holds $k_1 < \exp_f(\tau, P) \leq \lfloor \frac{\tau - s}{p} \rfloor$. Observe also that P = T[SA[i] ... n] implies that $SA[i] \in Occ(P, T)$. In particular, $Occ(P, T) \neq \emptyset$. Lastly, recall that type $(\tau, P) = -1$. By Lemma 8.110, we thus have that, letting $i' = \text{RangeBeg}_{\ell}(P, T) + \delta_{f, \ell}^{\text{low}-}(P, T)$, it holds $c + (i - i') \in [1 \dots \text{mod-count}_{\mathcal{I}, p}(s)]$ and $\exp_f(\tau, P) = \text{mod-select}_{\mathcal{I}, p}(s, c + (i - i'))$. By $i' = b + \delta$, we thus obtain $c + (i - (b + \delta)) \in [1 \dots \mathsf{mod\text{-}count}_{\mathcal{I},p}(s)]$ and $\exp_f(\tau, T, \mathrm{SA}[i]) = \exp_f(\tau, T[\mathrm{SA}[i] \dots n]) = \exp_f(\tau, P) = (i - (b + \delta))$ $\mathsf{mod}\text{-}\mathsf{select}_{\mathcal{I},p}(s,c+(i-(b+\delta))).$

Query Algorithms

Proposition 8.112. Let $k \in [4 ... \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $i \in [1 ... n]$ be such that $SA[i] \in R^-(\tau,T)$ and $e(\tau,T,SA[i]) - SA[i] < 2\ell$. Given CompSAPeriodic(T), the value k, the position i, some $j \in \text{Occ}_{3\tau-1}(\text{SA}[i], T)$ satisfying $j = \min \text{Occ}_{2\ell}(j, T)$, and the values $\text{head}_f(\tau, T, \text{SA}[i])$, $[\operatorname{root}_f(\tau, T, \operatorname{SA}[i])]$, Range $\operatorname{Beg}_\ell(\operatorname{SA}[i], T)$, $\operatorname{exp}_f^{\operatorname{cut}}(\tau, T, \operatorname{SA}[i], \ell)$, and $\delta_{f,\ell}^{\operatorname{low}}(\operatorname{SA}[i], T)$ as input, we can compute $\exp_f(\tau, T, SA[i])$ in $\mathcal{O}(\log^{3+\epsilon} n)$ time.

Proof. Denote $s = \text{head}_f(\tau, T, SA[i]), H = \text{root}_f(\tau, T, SA[i]), p = |H|, b = \text{RangeBeg}_\ell(SA[i], T), k_1 = \text{root}_\ell(T, T, SA[i])$ $\exp_f^{\text{cut}}(\tau, T, \text{SA}[i], \ell), \ \delta = \delta_{f,\ell}^{\text{low}}(\text{SA}[i], T), \ \text{and} \ \mathcal{I} = \text{WeightedIntervals}_{7\tau}(\text{Seed}_{f,H}^-(\tau, T), T). \ \text{First, we check}$ if $i \leq b + \delta$. If so, then by Lemma 8.111, it holds $\exp_f(\tau, T, SA[i]) = \exp_f^{\text{cut}}(\tau, T, SA[i], \ell)$, and thus we return k_1 as the output. Let us now assume $i > b + \delta$. Using Proposition 8.91 and the position j as input, in $\mathcal{O}(\log n)$ time we retrieve the pointer to the structure from Proposition 7.3 for Seed $_{tH}^{\tau}(\tau,T)$ $(j \in \mathsf{R}_{f,H}(\tau,T))$ follows by Lemma 8.50(2)), i.e., performing weighted modular constraint queries on $\mathcal{I}=$ WeightedIntervals_{7 τ}(Seed⁻_{f,H}(τ , T), T). Note that the pointer is not null, since we assumed SA[i] $\in R^-(\tau, T)$. Thus, $\mathsf{R}_{f,H}^-(\tau,T) \neq \emptyset$, which implies $\mathrm{Seed}_{f,H}^-(\tau,T) \neq \emptyset$. Next, using Proposition 7.3 in $\mathcal{O}(\log^{2+\epsilon}n)$ time we

compute $c = \mathsf{mod\text{-}count}_{\mathcal{I},p}(s,k_1)$. In $\mathcal{O}(\log^{3+\epsilon} n)$ time, we then compute $k = \mathsf{mod\text{-}select}_{\mathcal{I},p}(s,c+(i-(b+\delta)))$. By Lemma 8.111, it holds $k = \exp_f(\tau, T, SA[i])$. We thus return k as the answer. In total, we spend $\mathcal{O}(\log^{3+\epsilon} n)$ time.

8.4.10 Computing the Size of Occ

Combinatorial Properties

Lemma 8.113. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a $\tau\text{-periodic pattern such that }\operatorname{type}(\tau,P)=-1\ \ and\ T[n]\ \ does\ \ not\ \ occur\ \ in\ P[1\mathinner{\ldotp\ldotp} n).\ \ Let\ i\in\mathsf{R}(\tau,T).\ \ Denote\ H=\operatorname{root}_f(\tau,P)\ \ and\ t=e(\tau,T,i)-i-3\tau+2.\ \ Then,\ |\operatorname{Occ}_{2\ell}(P,T)\cap\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)\cap[i\mathinner{\ldotp\ldotp\ldotp} i+t)|\leq 1.\ \ Moreover,$ $|\operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \cap [i\ldots i+t)| = 1 \text{ holds if and only if}$

- $\begin{array}{l} \bullet \ i \in \mathsf{R}^-_{f,H}(\tau,T), \\ \bullet \ e^{\mathrm{full}}_f(\tau,T,i) i \geq e^{\mathrm{cut}}_f(\tau,P,2\ell) 1, \ and \\ \bullet \ P[e^{\mathrm{cut}}_f(\tau,P,2\ell) \ldots \min(m,2\ell)] \ is \ a \ prefix \ of \ T^\infty[e^{\mathrm{full}}_f(\tau,T,i) \ldots e^{\mathrm{full}}_f(\tau,T,i) + 7\tau). \end{array}$

Lastly, if $\operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \cap [i ... i+t) \neq \emptyset$, then $\operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \cap [i ... i+t) = \{e_f^{\operatorname{full}}(\tau,T,i) - (e_f^{\operatorname{cut}}(\tau,P,2\ell)-1)\}.$

Proof. Denote $s = \text{head}_f(\tau, P)$, p = |H|, and $k_2 = \exp_f^{\text{cut}}(\tau, P, 2\ell)$. By definition, $e(\tau, T, i) - i \geq 3\tau - 1$. Proof. Denote $s = \text{head}_f(\tau, P)$, p = |H|, and $k_2 = \exp_f^{\text{cat}}(\tau, P, 2\ell)$. By definition, $e(\tau, T, i) - i \geq 3\tau - 1$. Thus, t > 0. By Lemma 8.49(1), it holds $[i ... e(\tau, T, i) - 3\tau + 1] = [i ... i + t) \subseteq R(\tau, T)$. From Lemma 8.48, we thus obtain that for every $\delta \in [0 ... t)$, it holds $e_f^{\text{tull}}(\tau, T, i + \delta) = e_f^{\text{tull}}(\tau, T, i)$, which implies $e_f^{\text{tull}}(\tau, T, i + \delta) - (i + \delta) = e_f^{\text{tull}}(\tau, T, i) - i - \delta$. Consider any $j \in \text{Occ}_2(P, T) \cap \text{Pos}_{f,\ell}^{\text{high}}(P, T)$. By definition of $\text{Pos}_{f,\ell}^{\text{high}}(P, T)$, we then have $j \in R_{f,s,k_2,H}^-(\tau, T)$. Thus, $e_f^{\text{full}}(\tau, T, j) - j = s + k_2p = e_f^{\text{cut}}(\tau, P, 2\ell) - 1$. Consequently, $i + \delta \in \text{Occ}_2(P, T) \cap \text{Pos}_{f,\ell}^{\text{high}}(P, T)$ implies $e_f^{\text{tull}}(\tau, T, i + \delta) - (i + \delta) = e_f^{\text{tull}}(\tau, T, i) - i - \delta = e_f^{\text{cut}}(\tau, P, 2\ell) - 1$, i.e., $\delta = (e_f^{\text{tull}}(\tau, T, i) - i) - (e_f^{\text{cut}}(\tau, P, 2\ell) - 1)$, and hence $|\text{Occ}_2(P, T) \cap \text{Pos}_{f,\ell}^{\text{high}}(P, T) \cap |i ... i + t)| \leq 1$. We now prove the equivalence. Let us first assume that $|\text{Occ}_2(P, T) \cap \text{Pos}_{f,\ell}^{\text{high}}(P, T) \cap |i ... i + t)| = 1$, i.e., that for some $\delta \in [0 ... t)$, it holds $i + \delta \in \text{Occ}_2(P, T) \cap \text{Pos}_{f,\ell}^{\text{high}}(P, T)$. As noted above, this implies $i + \delta \in \mathbb{R}^r$, $u(\tau, T)$. By $[i ... i + \delta] \subseteq \mathbb{R}(\tau, T)$ and Lemma 8.48, this implies $i \in \mathbb{R}^r$, $u(\tau, T)$, i.e., the first condition.

 $\mathsf{R}^-_{f,s,k_2,H}(\tau,T)$. By $[i\mathinner{.\,.} i+\delta]\subseteq \mathsf{R}(\tau,T)$ and Lemma 8.48, this implies $i\in \mathsf{R}^-_{f,H}(\tau,T)$, i.e., the first condition. Second, recall from the first paragraph that $i+\delta\in \mathrm{Occ}_{2\ell}(P,T)\cap \mathrm{Pos}^{\mathrm{high}-}_{f,\ell}(P,T)$ implies $e^{\mathrm{cut}}_f(\tau,P,2\ell)-1=0$ $e_f^{\text{full}}(\tau, T, i) - i - \delta \leq e_f^{\text{full}}(\tau, T, i) - i$. This establishes the second condition. We now show the third condition. As noted above, $i + \delta \in \operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P,T)$ implies that $i + \delta \in \operatorname{R}_{f,s,k_2,H}(\tau,T)$. Consequently, $T[i + \delta \dots e_f^{\operatorname{full}}(\tau,T,i+\delta)) = T[i + \delta \dots e_f^{\operatorname{full}}(\tau,T,i)) = P[1 \dots e_f^{\operatorname{cut}}(\tau,P,2\ell)) = H'H^{k_2}$, where H' is a length-ssuffix of H. Thus, $P[1..\min(m, 2\ell)]$ being a prefix of $T[i + \delta..n]$ (which follows from $i + \delta \in \text{Occ}_{2\ell}(P, T)$) implies that $P[e_f^{\text{cut}}(\tau, P, 2\ell) \dots \min(m, 2\ell)]$ is a prefix of $T[e_f^{\text{full}}(\tau, T, i) \dots n]$. By $|P[e_f^{\text{cut}}(\tau, P, 2\ell) \dots \min(m, 2\ell)]| = \min(m, 2\ell) - e_f^{\text{cut}}(\tau, P, 2\ell) + 1 \le 2\ell \le 7\tau$ (where the last inequality follows by $\tau = \lfloor \frac{\ell}{3} \rfloor$ and $\ell \ge 16$), we thus obtain that $P[e_f^{\text{cut}}(\tau, P, 2\ell) \dots \min(m, 2\ell)]$ is a prefix of $T^{\infty}[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + 7\tau)$.

We now prove the opposite implication. Let $i \in R(\tau, T)$ and assume $i \in R_{f,H}^{-}(\tau, T)$, $e_f^{\text{full}}(\tau, T, i) - i \ge e_f^{\text{cut}}(\tau, P, 2\ell) - 1$, and $P[e_f^{\text{cut}}(\tau, P, 2\ell) \dots \min(m, 2\ell)]$ is a prefix of $T^{\infty}[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + 7\tau)$. Observe that this implies that $P[e_f^{\text{cut}}(\tau, P, 2\ell) \dots \min(m, 2\ell)] \le T^{\infty}[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + 7\tau)$ and hence by Lemma 8.92, letting $\delta = (e_f^{\text{full}}(\tau, T, i) - i) - (e_f^{\text{cut}}(\tau, P, 2\ell) - 1)$, we have $i + \delta \in [0 \dots t)$ and $i + \delta \in \text{Pos}_{f,\ell}^{\text{high}}(P, T)$. It thus remains to prove that $i + \delta \in \text{Occ}_{2\ell}(P, T)$, i.e., that $\text{lcp}(P, T[i + \delta ... n]) \ge \min(m, 2\ell)$, or equivalently, that $i + \delta + \min(m, 2\ell) - 1 \le n$ and $T[i + \delta ... i + \delta + \min(m, 2\ell) - 1] = P[1...\min(m, 2\ell)]$. We proceed in two steps:

• First, we prove that $T[i+\delta ...e_f^{\text{full}}(\tau,T,i)) = P[1...e_f^{\text{cut}}(\tau,P,2\ell))$. First, note that $[i...i+\delta] \subseteq \mathbb{R}(\tau,T)$, $i \in \mathbb{R}_{f,H}(\tau,T)$ and Lemma 8.48 imply that $i+\delta \in \mathbb{R}_{f,H}(\tau,T)$ and $e_f^{\text{full}}(\tau,T,i+\delta) = e_f^{\text{full}}(\tau,T,i)$. On the other hand, by definition of δ , we have $e_f^{\text{full}}(\tau,T,i) - (i+\delta) = e_f^{\text{cut}}(\tau,P,2\ell) - 1$. Consequently, $\text{head}_f(\tau,T,i+\delta) = (e_f^{\text{full}}(\tau,T,i+\delta) - (i+\delta)) \mod p = (e_f^{\text{cut}}(\tau,P,2\ell) - 1) \mod p = \text{head}_f(\tau,P) = s$. We also have $\exp_f(\tau,T,i+\delta) = \lfloor \frac{e_f^{\text{full}}(\tau,T,i+\delta) - (i+\delta)}{p} \rfloor = \lfloor \frac{e_f^{\text{cut}}(\tau,P,2\ell) - 1}{p} \rfloor = \lfloor$ suffix of H. On the other hand, we also have by definition $P[1..e_f^{\text{cut}}(\tau,P,2\ell)) = H'H^{k_2}$. Thus, $T[i + \delta \dots e_f^{\text{full}}(\tau, T, i)) = P[1 \dots e_f^{\text{cut}}(\tau, P, 2\ell)).$

• Second, we prove that it holds $i + \delta + \min(m, 2\ell) - 1 \le n$ and $T[e_f^{\text{full}}(\tau, T, i) ... i + \delta + \min(m, 2\ell) - 1] = 0$ $P[e_f^{\text{cut}}(\tau, P, 2\ell) \dots \min(m, 2\ell)]$. Recall, that we assumed that $P[e_f^{\text{cut}}(\tau, P, 2\ell) \dots \min(m, 2\ell)]$ is a prefix of $T^{\infty}[e_f^{\text{full}}(\tau, T, i) \dots e_f^{\text{full}}(\tau, T, i) + 7\tau)$. Note, however, that we also assumed that T[n] does not occur in P[1..m). Consequently, T[n] also does not occur in $P[e_f^{\text{cut}}(\tau, P, 2\ell)..\min(m, 2\ell))$, and thus, $e_f^{\text{full}}(\tau,T,i) + (\min(m,2\ell) - e_f^{\text{cut}}(\tau,P,2\ell)) \leq n. \text{ By recalling that } i + \delta = i + (e_f^{\text{full}}(\tau,T,i) - i) - (e_f^{\text{cut}}(\tau,P,2\ell) - 1) = e_f^{\text{full}}(\tau,T,i) - e_f^{\text{cut}}(\tau,P,2\ell) + 1, \text{ we equivalently obtain } e_f^{\text{full}}(\tau,T,i) + \min(m,2\ell) - e_f^{\text{cut}}(\tau,P,2\ell) = e_f^{\text{full}}(\tau,T,i) + e_f^{\text{fu$ $e_f^{\text{cut}}(\tau, P, 2\ell) = i + \delta + \min(m, 2\ell) - 1 \leq n$, i.e., the first part of the claim. Moreover, observe that $P[e_f^{\text{cut}}(\tau,P,2\ell)\ldots\min(m,2\ell)]$ being a prefix of $T[e_f^{\text{full}}(\tau,T,i)\ldots e_f^{\text{full}}(\tau,T,i)+7\tau)$ then implies that it holds $P[e_f^{\text{cut}}(\tau,P,2\ell)\ldots\min(m,2\ell)] = T[e_f^{\text{full}}(\tau,T,i)\ldots e_f^{\text{full}}(\tau,T,i)+\min(m,2\ell)-e_f^{\text{cut}}(\tau,P,2\ell)] = T[e_f^{\text{full}}(\tau,T,i)\ldots i+\delta+\min(m,2\ell)-1]$, i.e., the second part of the claim.

To show the last implication, recall that we proved that $i+\delta\in \mathrm{Occ}_{2\ell}(P,T)\cap \mathrm{Pos}^{\mathrm{high-}}_{f,\ell}(P,T)$ (where $\delta\in[0\mathinner{\ldotp\ldotp} t)$) implies $\delta=(e^{\mathrm{full}}_f(\tau,T,i)-i)-(e^{\mathrm{cut}}_f(\tau,P,2\ell)-1)$. Thus, $\mathrm{Occ}_{2\ell}(P,T)\cap \mathrm{Pos}^{\mathrm{high-}}_{f,\ell}(P,T)\cap[i\mathinner{\ldotp\ldotp} i+t)\neq\emptyset$ implies $\mathrm{Occ}_{2\ell}(P,T)\cap \mathrm{Pos}^{\mathrm{high-}}_{f,\ell}(P,T)\cap[i\mathinner{\ldotp\ldotp} i+t)=\{i+\delta\}=\{e^{\mathrm{full}}_f(\tau,T,i)-(e^{\mathrm{cut}}_f(\tau,P,2\ell)-1)\}$. \square

Lemma 8.114. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{type}(\tau, P) = -1$ and T[n] does not occur in P[1..m). Denote $H = \operatorname{root}_f(\tau, P)$, $s = \text{head}_f(\tau, P), \text{ and } k_2 = \exp_f^{\text{cut}}(\tau, P, 2\ell). \text{ For every } i \in \mathsf{R}(\tau, T), i \in \operatorname{Occ}_{2\ell}(P, T) \cap \operatorname{Pos}_{f,\ell}^{\text{high}}(P, T) \text{ holds if } i$ and only if

- $\begin{array}{l} \bullet \ i \in \mathsf{R}^-_{f,s,k_2,H}(\tau,T) \ and \\ \bullet \ P[e^{\mathrm{cut}}_f(\tau,P,2\ell) \ldots \min(m,2\ell)] \ is \ a \ prefix \ of \ T^\infty[e^{\mathrm{full}}_f(\tau,T,i) \ldots e^{\mathrm{full}}_f(\tau,T,i) + 7\tau). \end{array}$

Proof. Consider any $i \in \mathsf{R}(\tau,T)$ and assume $i \in \mathsf{Occ}_{2\ell}(P,T) \cap \mathsf{Pos}_{f,\ell}^{\mathsf{high}-}(P,T)$. By Definition 8.70, this implies $i \in \mathsf{R}_{f,s,k_2,H}^-(\tau,T)$. To show the second claim, note that by $e(\tau,T,i)-i \geq 3\tau-1$, letting $t=e(\tau,T,i)-i-3\tau+2$, we have $\mathsf{Occ}_{2\ell}(P,T) \cap \mathsf{Pos}_{f,\ell}^{\mathsf{high}-}(P,T) \cap [i\ldots i+t) \neq \emptyset$. By Lemma 8.113, we thus obtain that $P[e_f^{\mathsf{cut}}(\tau,P,2\ell)\ldots \mathsf{min}(m,2\ell)]$ is a prefix of $T^\infty[e_f^{\mathsf{full}}(\tau,T,i)\ldots e_f^{\mathsf{full}}(\tau,T,i)+7\tau)$. Let us now consider $i \in \mathsf{R}(\tau,T)$ and assume $i \in \mathsf{R}_{f,s,k_2,H}^-(\tau,T)$ and that $P[e_f^{\mathsf{cut}}(\tau,P,2\ell)\ldots \mathsf{min}(m,2\ell)]$ is a prefix of $T^\infty[e_f^{\mathsf{full}}(\tau,T,i)\ldots e_f^{\mathsf{full}}(\tau,T,i)+7\tau)$. The first assumption implies $e_f^{\mathsf{full}}(\tau,T,i)-i=s+k_2\cdot |H|=\mathsf{head}_f(\tau,P)+\mathsf{exp}_f^{\mathsf{cut}}(\tau,P,2\ell)\cdot |\mathsf{root}_f(\tau,P)|=e_f^{\mathsf{cut}}(\tau,P,2\ell)-1$. By Lemma 8.113, it thus follows that, letting $t=e(\tau,T,i)-i-3\tau+2>0$, we have $\mathsf{Occ}_{2\ell}(P,T)\cap \mathsf{Pos}_{f,\ell}^{\mathsf{high}-}(P,T)\cap [i\ldots i+t)=\{e_f^{\mathsf{full}}(\tau,T,i)-(e_f^{\mathsf{cut}}(\tau,P,2\ell)-1)\}=\{i\}$. Thus, $i\in \mathsf{Occ}_{2\ell}(P,T)\cap \mathsf{Pos}_{f,\ell}^{\mathsf{high}-}(P,T)$.

Lemma 8.115. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{type}(\tau, P) = -1$ and T[n] does not occur in P[1..m). Let $c = \max \Sigma$, $x_l = e_f^{\text{cut}}(\tau, P, 2\ell) - 1$, $y_l = P[e_f^{\text{cut}}(\tau, P, 2\ell) \dots \min(m, 2\ell)]$, and $y_u = y_l c^{\infty}$. Then, letting $H = \text{root}_f(\tau, P)$ and $\mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f,H}^-(\tau,T),T), it holds:$

- $$\begin{split} \bullet & \operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) = \{e_f^{\operatorname{full}}(\tau,T,j) x_l : j \in \mathsf{R}_{f,H}^{\prime-}(\tau,T), \\ & x_l \leq e_f^{\operatorname{full}}(\tau,T,j) j, \ and \ y_l \preceq T^{\infty}[e_f^{\operatorname{full}}(\tau,T,j) \ldots e_f^{\operatorname{full}}(\tau,T,j) + 7\tau) \prec y_u\}, \\ \bullet & | \operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)| = \operatorname{weight-count}_{\mathcal{P}}(x_l,n,y_l,y_u). \end{split}$$

Proof. By Definition 8.70, we have $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)\subseteq \mathsf{R}_{f,H}^{-}(\tau,T)$. On the other hand, by Lemma 8.49(1) it holds $\mathsf{R}_{f,H}^{-}(\tau,T)=\bigcup_{j\in\mathsf{R}_{f,H}^{-}(\tau,T)}[j\cdot\cdot e(\tau,T,j)-3\tau+1]$. Since this is a disjoint union, we thus have $\operatorname{Occ}_{2\ell}(P,T)\cap\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)=\bigcup_{j\in\mathsf{R}_{f,H}^{\prime}(\tau,T)}[j\cdot\cdot e(\tau,T,j)-3\tau+1]$. By Lemma 8.113, for every $j\in\mathsf{R}_{f,H}^{\prime}(\tau,T), |\operatorname{Occ}_{2\ell}(P,T)\cap\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)\cap[j\cdot\cdot e(\tau,T,j)-3\tau+1]|\leq 1$. Moreover, $|\operatorname{Occ}_{2\ell}(P,T)\cap\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)\cap[j\cdot\cdot e(\tau,T,j)-3\tau+1]|=1$ holds if and only if $x_l\leq e_f^{\operatorname{full}}(\tau,T,j)-j$ and y_l is a prefix of $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,j)\cdot\cdot e_f^{\operatorname{full}}(\tau,T,j)+7\tau)$. Furthermore, if $x_l\leq e_f^{\operatorname{full}}(\tau,T,j)-j$ and y_l is a prefix of $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,j)\cdot\cdot e_f^{\operatorname{full}}(\tau,T,j)+7\tau)$, then $\operatorname{Occ}_{2\ell}(P,T)\cap\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)\cap[j\cdot\cdot e(\tau,T,j)-3\tau+1]=\{e_f^{\operatorname{full}}(\tau,T,j)-x_l\}$. Consequently, letting

$$\mathcal{J} = \{ j \in \mathsf{R}_{f,H}^{\prime-}(\tau,T) : x_l \leq e_f^{\mathrm{full}}(\tau,T,j) - j \text{ and } y_u \text{ is a prefix of } T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j) \dots e_f^{\mathrm{full}}(\tau,T,j) + 7\tau) \}$$

$$= \{ j \in \mathsf{R}_{f,H}^{\prime-}(\tau,T) : x_l \leq e_f^{\mathrm{full}}(\tau,T,j) - j \text{ and } y_l \leq T^{\infty}[e_f^{\mathrm{full}}(\tau,T,j) \dots e_f^{\mathrm{full}}(\tau,T,j) + 7\tau) \prec y_u \},$$

we have $\operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) = \bigcup_{j \in \mathcal{J}} \{e_f^{\operatorname{full}}(\tau,T,j) - x_l\}$, i.e., the first claim.

To show the second claim, note that for every $j \in \mathcal{J}$, we have $e_f^{\text{full}}(\tau, T, j) - x_l \in [j ... e(\tau, T, j) - 3\tau + 1]$. Since by Lemma 8.49(1), for every $j_1, j_2 \in R'(\tau, T), j_1 \neq j_2$ implies that $[j_1 ... e(\tau, T, j_1) - 3\tau + 1]$ and $[j_2 ... e(\tau, T, j_2) - 1]$ $3\tau+1$] are disjoint, we thus obtain that $|\operatorname{Occ}_{2\ell}(P,T)\cap\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)|=|\bigcup_{j\in\mathcal{J}}\{e_{\mathrm{full}}^{\mathrm{full}}(\tau,T,j)-x_l\}|=|\mathcal{J}|.$ Consequently, it follows by Lemma 8.81(1) that $|\operatorname{Occ}_{2\ell}(P,T)\cap\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)|=\operatorname{weight-count}_{\mathcal{P}}(x_l,n,y_l,y_u).$ Note that Lemma 8.81(1) requires $x_l\in[0...7\tau]$, which holds here since $x_l=e_f^{\mathrm{cut}}(\tau,P,2\ell)-1\leq 2\ell\leq 7\tau$, where the last inequality holds by $\tau = \lfloor \frac{\ell}{3} \rfloor$ and $\ell \geq 16$.

Lemma 8.116. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{type}(\tau, P) = -1$, $e(\tau, P) - 1 < \min(m, 2\ell)$, and T[n] does not occur in P[1..m). Then, it holds $\exp_f(\tau, P) = \exp_f^{\text{cut}}(\tau, P, 2\ell)$, $e_f^{\text{full}}(\tau, P) = e_f^{\text{cut}}(\tau, P, 2\ell)$, and $\operatorname{Occ}_{2\ell}(P, T) \subseteq \operatorname{Pos}_{f\ell}^{\text{high}}(P, T)$.

Proof. Denote $s = \text{head}_f(\tau, P), \ p = |H|, \ k = \exp_f(\tau, P), \ \text{and} \ k_2 = \exp_{f}^{\text{cut}}(\tau, P, 2\ell).$ We start by observing that $e(\tau, P) - 1 < 2\ell$ implies that $k_2 = \min(\exp_f(\tau, P), \lfloor \frac{2\ell - s}{p} \rfloor) = \min(\lfloor \frac{e(\tau, P) - 1 - s}{p} \rfloor, \lfloor \frac{2\ell - s}{p} \rfloor) = \lfloor \frac{e(\tau, P) - 1 - s}{p} \rfloor = \exp_f(\tau, P) = k$. This implies $e_f^{\text{full}}(\tau, P) = 1 + s + kp = 1 + s + kp = e_f^{\text{cut}}(\tau, P, 2\ell) = e_f^{\text{cut}}(\tau, P, 2\ell).$ Next, we prove $\text{Occ}_2(P, T) \subseteq \text{Pos}_{f,\ell}^{\text{high}-}(P, T).$ Let $t \in \text{Occ}_2(P, T).$ By Lemma 8.117, it follows that $t \in \mathbb{R}_{f,s,k,H}^-(\tau, T) = \mathbb{R}_{f,s,k_2,H}^-(\tau, T)$ and, letting $P_{\text{sub}} = P[e_f^{\text{full}}(\tau, P) \dots \min(m, 2\ell)] = P[e_f^{\text{cut}}(\tau, P, 2\ell) \dots \min(m, 2\ell)],$ the string P_{sub} is a prefix of $T^{\infty}[e_f^{\text{full}}(\tau, T, t) \dots e_f^{\text{full}}(\tau, T, t) + 7\tau).$ Thus, $T^{\infty}[e_f^{\text{full}}(\tau, T, t) \dots e_f^{\text{full}}(\tau, T, t) + 7\tau) \succeq P_{\text{sub}}.$ By Lemma 8.93, we thus have $t \in \text{Pos}_{f,\ell}^{\text{high}-}(P, T).$

Lemma 8.117. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{type}(\tau, P) = -1$, $e(\tau, P) - 1 < \min(m, 2\ell)$, and T[n] does not occur in P[1 ...m). Denote $H = \operatorname{root}_f(\tau, P)$, $s = \operatorname{head}_f(\tau, P)$, and $k = \exp_f(\tau, P)$. For every $i \in R(\tau, T)$, $i \in \operatorname{Occ}_{2\ell}(P, T)$ holds if and only if

- $\begin{array}{l} \bullet \ i \in \mathsf{R}^-_{f,s,k,H}(\tau,T) \ \ and \\ \bullet \ \ P[e^{\mathrm{full}}_f(\tau,P) \ldots \min(m,2\ell)] \ \ is \ a \ \ prefix \ of \ T^\infty[e^{\mathrm{full}}_f(\tau,T,i) \ldots e^{\mathrm{full}}_f(\tau,T,i) + 7\tau). \end{array}$

Proof. Let us denote $k_2 = \exp_f^{\text{cut}}(\tau, P, 2\ell)$. By Lemma 8.116, it holds $k_2 = k$, $e_f^{\text{cut}}(\tau, P, 2\ell) = e_f^{\text{full}}(\tau, P)$, and $\operatorname{Occ}_{2\ell}(P,T) \subseteq \operatorname{Pos}_{f,\ell}^{\text{high}-}(P,T)$. Thus, it follows by Lemma 8.114 that for every $i \in \mathsf{R}(\tau,T)$, $i \in \operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\text{high}-}(P,T) = \operatorname{Occ}_{2\ell}(P,T)$ holds if and only if $i \in \mathsf{R}_{f,s,k_2,H}^-(\tau,T) = \mathsf{R}_{f,s,k,H}^-(\tau,T)$ and $P[e_f^{\text{cut}}(\tau,P,2\ell) \dots \min(m,2\ell)] = P[e_f^{\text{full}}(\tau,P) \dots \min(m,2\ell)]$ is a prefix of $T^{\infty}[e_f^{\text{full}}(\tau,T,i) \dots e_f^{\text{full}}(\tau,T,i) + 7\tau)$.

Lemma 8.118. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{type}(\tau,P) = -1$, $e(\tau,P) - 1 < \min(m,2\ell)$, and T[n] does not occur in P[1..m). Let $c = \max \Sigma$, $x_l = e_f^{\text{full}}(\tau, P) - 1$, $y_l = P[e_f^{\text{full}}(\tau, P) ... \min(m, 2\ell)]$, and $y_u = y_l c^{\infty}$. Then, letting $H = \text{root}_f(\tau, P)$ and $\mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f, H}^-(\tau, T), T)$, it holds:

- $\operatorname{Occ}_{2\ell}(P,T) = \{e_f^{\text{full}}(\tau,T,j) x_l : j \in \mathsf{R}_{f,H}^{\prime-}(\tau,T), \ x_l \leq e_f^{\text{full}}(\tau,T,j) j, \ and \ y_l \leq T^{\infty}_{f,H}[e_f^{\text{full}}(\tau,T,j) ... e_f^{\text{full}}(\tau,T,j) + 7\tau) \prec y_u\},$
- $|\operatorname{Occ}_{2\ell}(P, T)| = \operatorname{weight-count}_{\mathcal{P}}(x_l, n, y_l, y_u).$

 $\begin{array}{l} \textit{Proof.} \text{ By Lemma 8.116, it holds } e_f^{\text{cut}}(\tau,P,2\ell)-1=e_f^{\text{full}}(\tau,P)-1=x_l, \text{ and } \operatorname{Occ}_{2\ell}(P,T)\subseteq\operatorname{Pos}_{f,\ell}^{\text{high}-}(P,T). \\ \text{Observe that we then also have } P[e_f^{\text{cut}}(\tau,P,2\ell)\ldots\min(m,2\ell)]=P[e_f^{\text{full}}(\tau,P)\ldots\min(m,2\ell)]=y_l. \text{ Thus, } \\ \text{it follows by Lemma 8.115 that } \operatorname{Occ}_{2\ell}(P,T)\cap\operatorname{Pos}_{f,\ell}^{\text{high}-}(P,T)=\operatorname{Occ}_{2\ell}(P,T)=\{e_f^{\text{full}}(\tau,T,j)-x_l:j\in R_{f,H}^{\prime}(\tau,T),\ x_l\leq e_f^{\text{full}}(\tau,T,j)-j, \text{ and } y_l\preceq T^{\infty}[e_f^{\text{full}}(\tau,T,j)\ldots e_f^{\text{full}}(\tau,T,j)+7\tau)\prec y_u\} \text{ and } |\operatorname{Occ}_{2\ell}(P,T)\cap\operatorname{Pos}_{f,\ell}^{\text{high}-}(P,T)|=|\operatorname{Occ}_{2\ell}(P,T)|=\text{weight-count}_{\mathcal{P}}(x_l,n,y_l,y_u). \\ \end{array}$

Lemma 8.119. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $j \in \mathbb{R}^-(\tau,T)$ be such that $e(\tau, T, j) - j < 2\ell$. Let $c = \max \Sigma$, $x_l = e_f^{\text{full}}(\tau, T, j) - j$, $y_l = T[e_f^{\text{full}}(\tau, T, j) \dots \min(n+1, j+1)]$ (2ℓ)), and $y_u = y_l c^{\infty}$. Then, letting $H = \text{root}_f(\tau, T, j)$ and $\mathcal{P} = \text{IntStrPoints}_{\tau}(\text{Seed}_{f,H}^-(\tau, T), T)$, it holds $|\operatorname{Occ}_{2\ell}(j,T)| = \mathsf{weight\text{-}count}_{\mathcal{P}}(x_l,n,y_l,y_u).$

Proof. Denote P = T[j ... n] and m = |P|. First, observe that by $j \in R^-(\tau, T)$, P is τ -periodic and it holds type $(\tau, P) = -1$. Note also that T[n] does not occur in P[1..m). Next, recall that, by definition, $e(\tau,T,j)=j+e(\tau,P)-1$. Thus, we have $e(\tau,P)-1=e(\tau,T,j)-j<2\ell$. Since by the uniqueness of T[n] in T it holds $e(\tau, T, j) \leq n$, we also have $e(\tau, P) - 1 = e(\tau, T, j) - j \leq n - j < n - j + 1 = m$. We have thus proved that $e(\tau, P) - 1 < \min(m, 2\ell)$. Next, observe that $e_f^{\text{full}}(\tau, T, j) = j + e_f^{\text{full}}(\tau, P) - 1$. Thus, $e_f^{\text{full}}(\tau, P) - 1 = e_f^{\text{full}}(\tau, T, j) - j = x_l$. Moreover, by P = T[j ... n],

$$\begin{split} P[e_f^{\text{full}}(\tau, P) \ldots & \min(m, d)] = T[j + e_f^{\text{full}}(\tau, P) - 1 \ldots j + \min(m, 2\ell) - 1] \\ &= T[e_f^{\text{full}}(\tau, T, j) \ldots (j - 1) + \min(n - (j - 1), 2\ell)] \\ &= T[e_f^{\text{full}}(\tau, T, j) \ldots \min(n, j + 2\ell - 1)] \\ &= T[e_f^{\text{full}}(\tau, T, j) \ldots \min(n + 1, j + 2\ell)) \\ &= y_l. \end{split}$$

Lastly, by definition we have $\operatorname{root}_f(\tau, T, j) = \operatorname{root}_f(\tau, T[j \dots n]) = \operatorname{root}_f(\tau, P)$, and hence $\operatorname{root}_f(\tau, P) = H$. By Lemma 8.118, we therefore obtain that $|\operatorname{Occ}_{2\ell}(P,T)| = \operatorname{weight-count}_{\mathcal{P}}(x_l,n,y_l,y_u)$. Since by definition we have $\operatorname{Occ}_{2\ell}(j,T) = \operatorname{Occ}_{2\ell}(T[j..n],T) = \operatorname{Occ}_{2\ell}(P,T)$, we thus obtain the claim.

Query Algorithms

Proposition 8.120. Let $k \in [4 ... \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $j \in \mathbb{R}^-(\tau,T)$ be such that $e(\tau, T, j) - j < 2\ell$. Given CompSAPeriodic(T), the value k, the position j, some $j' \in \text{Occ}_{3\tau-1}(j, T)$ satisfying $j' = \min \operatorname{Occ}_{2\ell}(j',T)$, and the values $\operatorname{head}_f(\tau,T,j)$, $|\operatorname{root}_f(\tau,T,j)|$, and $\exp_f(\tau,T,j)$ as input, we can compute $|\operatorname{Occ}_{2\ell}(j,T)|$ in $\mathcal{O}(\log^{2+\epsilon} n)$ time.

 $\textit{Proof.} \text{ Let us denote } s = \text{head}_f(\tau, T, j), \ H = \text{root}_f(\tau, T, j), \ p = |H|, \ k = \exp_f(\tau, T, j), \ c = \max \Sigma,$ $y_l = T[e_f^{\text{full}}(\tau, T, j) \dots \min(n+1, j+2\ell)), y_u = y_l c^{\infty}, \text{ and } \mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f,H}^{-}(\tau, T), T). \text{ In } \mathcal{O}(1) \text{ time we set } x_l := e_f^{\text{full}}(\tau, T, j) - j = s + kp. \text{ Using Proposition 8.90 and the position } j' \text{ as input, in } \mathcal{O}(\log n) \text{ time } j' \text{ as input, in } \mathcal{O}(\log n) \text{ time } j' \text{ as input, in } \mathcal{O}(\log n) \text{ time } j' \text{ the proposition 8.90}$ we retrieve the pointer to the structure from Proposition 6.6 for Seed $_{f,H}^-(\tau,T)$ (note that $j' \in \mathsf{R}_{f,H}(\tau,T)$) holds by Lemma 8.50(2)), i.e., performing weighted range queries on $\mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f,H}^{-}(\tau,T),T)$. Note that this pointer is not null, since we assumed $j \in \mathsf{R}^-(\tau,T)$, which implies $\mathrm{Seed}_{f,H}^-(\tau,T) \neq \emptyset$. Note that using CompSACore(T) (which is part of CompSAPeriodic(T)), we can lexicographically compare any two substrings of T^{∞} or $\overline{T^{\infty}}$ (specified with their starting positions and lengths) in $t_{\text{cmp}} = \mathcal{O}(\log n)$ time. In $\mathcal{O}(\log^{2+\epsilon} n + t_{\text{cmp}} \log n) = \mathcal{O}(\log^{2+\epsilon} n)$ time we therefore compute $q := \text{weight-count}_{\mathcal{P}}(x_l, n, y_l, y_u) = 0$ weight-count_P (x_l, n, y_u) - weight-count_P (x_l, n, y_l) using Proposition 6.6 with $i = j + x_l$ and $q_r = \min(n + i)$ $1, j + 2\ell$) – i. By Lemma 8.119, it holds $q = |\operatorname{Occ}_{2\ell}(j,T)|$. In total, we spend $\mathcal{O}(\log^{2+\epsilon} n)$ time.

8.4.11 Computing a Position in Occ

Combinatorial Properties

Lemma 8.121. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{type}(\tau, P) = -1$ and T[n] does not occur in P[1..m). Denote $H = \operatorname{root}_f(\tau, P)$, $\mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f,H}^{-}(\tau,T),T), \ i' = \operatorname{RangeBeg}_{\ell}(P,T) + \delta_{f,\ell}^{\operatorname{low}-}(P,T) + \delta_{f,\ell}^{\operatorname{mid}-}(P,T), \ x = e_f^{\operatorname{cut}}(\tau,P,2\ell) - 1, \\ and \ c = \operatorname{weight-count}_{\mathcal{P}}(x,n). \ \ Let \ i \in [1 \ldots n] \ \ be \ such \ that \ \operatorname{SA}[i] \in \operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T). \ \ Then:$

- It holds $c (i' i) \in [1 ... c]$, i.e., weight-select_P(x, n, c (i' i)) is well-defined,
- Every position $j \in \text{weight-select}_{\mathcal{P}}(x, n, c (i' i))$ satisfies $j x \in \text{Occ}_{2\ell}(P, T) \cap \text{Pos}_{f, \ell}^{\text{high-}}(P, T)$.

Proof. Denote $s = \text{head}_f(\tau, P)$, p = |H|, and $k_2 = \exp_f^{\text{cut}}(\tau, P, 2\ell)$. The proof consists of three steps.

1. First, we prove $c - (i' - i) \in [1 \dots c]$, i.e., the first claim. Recall that by Lemma 8.107, it holds $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) = \{\operatorname{SA}[i''] : i'' \in (\operatorname{RangeBeg}_{2\ell}(P,T) \dots \operatorname{RangeBeg}_{2\ell}(P,T) + \delta_{f,\ell}^{\operatorname{high}-}(P,T)]\}$, which due to Lemma 8.74 and Lemma 8.6, we can rewrite as $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) = \{\operatorname{SA}[i''] : i'' \in (i' - \delta_{f,\ell}^{\operatorname{high}-}(P,T) \dots i']\}$. Consequently, $\operatorname{SA}[i] \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$ implies $i' - \delta_{f,\ell}^{\operatorname{high}-}(P,T) < i \le i'$. On the one hand, we thus obtain $i' - i < \delta_{f,\ell}^{\operatorname{high}-}(P,T) = |\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)| \le |\operatorname{R}_{f,s,k_2,H}^-(T,T)| \le |\operatorname{R}_{f,s,k_2,H}^-(T,T)|$ follows by $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \subseteq \operatorname{R}_{f,s,k_2,H}^-(T,T)$ and $|\operatorname{R}_{f,s,k_2,H}^-(T,T)| \le \operatorname{R}_{f,s,k_2,H}^-(T,T)$ follows by Lemma 8.82 Note that Lemma 8.82 $\mathsf{R}^-_{f,s,k_2,H}(\tau,T)$ and $|\mathsf{R}^-_{f,s,k_2,H}(\tau,T)| \leq \mathsf{weight\text{-}count}_{\mathcal{P}}(s+k_2p,n)$ follows by Lemma 8.82. Note that Lemma 8.82

- requires $k_2 \leq \lfloor \frac{7\tau s}{p} \rfloor$, which holds here since $k_2 = \min(\exp_f(\tau, P), \lfloor \frac{2\ell s}{p} \rfloor) \leq \lfloor \frac{2\ell s}{p} \rfloor \leq \lfloor \frac{7\tau s}{p} \rfloor$, where $2\ell \leq 7\tau$ follows by $\tau = \lfloor \frac{\ell}{3} \rfloor$ and $\ell \geq 16$. We thus have i' i < c, which is equivalent to $c (i' i) \geq 1$. On the other hand, above we proved that $i \leq i'$, which is equivalent to $c (i' i) \leq c$. We thus obtain $c (i' i) \in [1...c]$.
- hand, above we proved that $i \leq i'$, which is equivalent to $c (i' i) \leq c$. We thus obtain $c (i' i) \in [1 \dots c]$. Next, we show that for every $\delta \in [0 \dots \delta_{f,\ell}^{high-}(P,T))$, any $j \in \text{weight-select}_{\mathcal{P}}(x,n,c-\delta)$ satisfies $j-x \in [1 \dots n]$ and $T^{\infty}[j-x\dots j-x+7\tau) = T^{\infty}[\text{SA}[i'-\delta]\dots \text{SA}[i'-\delta]+7\tau)$. Consider any $\delta \in [0 \dots \delta_{f,\ell}^{high-}(P,T))$ and let $j' = \text{SA}[i'-\delta]$. Denote j'' = j' + x and let $y_u = T^{\infty}[j'' \dots j'' + 7\tau)$. We first prove that it holds $c \delta \in (\text{weight-count}_{\mathcal{P}}(x,n,y_u) \dots \text{weight-count}_{\mathcal{P}}^{\infty}(x,n,y_u)]$.
 - (a) First, we show that it holds $c-\delta > \text{weight-count}_{\mathcal{P}}(x,n,y_u)$. We start by observing that by Lemma 8.81(1), we have weight-count $_{\mathcal{P}}(x,n,y_u) = \text{weight-count}_{\mathcal{P}}(x,n,\varepsilon,y_u) = |A|$, where $A = \{t \in \mathsf{R}'_{f,H}(\tau,T) : x \leq e_f^{\mathrm{full}}(\tau,T,t) t \text{ and } T^\infty[e_f^{\mathrm{full}}(\tau,T,t) ... e_f^{\mathrm{full}}(\tau,T,t) + 7\tau) \prec y_u\}$. Note that Lemma 8.81 requires $x \leq 7\tau$, which holds here by $x = e_f^{\mathrm{cut}}(\tau,P,2\ell) 1 = s + k_2p = s + \min(\exp_f(\tau,P),\lfloor\frac{2\ell-s}{p}\rfloor)p \leq s + \lfloor\frac{2\ell-s}{p}\rfloor p \leq 2\ell \leq 7\tau$. On the other hand, by Lemma 8.81(3), we have c = |A'|, where $A' = \{t \in \mathsf{R}'_{f,H}(\tau,T) : x \leq e_f^{\mathrm{full}}(\tau,T,t) t\}$. Our claim is thus equivalent to $|A'| |A| > \delta$. Observe that $A \subseteq A'$. This implies $|A'| |A| = |A' \setminus A|$. Consequently, our goal is to prove that $|A' \setminus A| > \delta$. To this end, we will show that, letting $B = \{b(\tau,T,\mathrm{SA}[i'']) : i'' \in [i' \delta ... i']\}$, it holds $|B| = \delta + 1$ and $B \subseteq A' \setminus A$. We proceed in two steps:
 - First, we show that $|B| = \delta + 1$. Let $q, q' \in [i' \delta ...i']$ and assume $q \neq q'$. We will prove that $b(\tau, T, SA[q]) \neq b(\tau, T, SA[q'])$. Above we proved that $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) = \{SA[i''] : i'' \in (i' \delta_{f,\ell}^{\operatorname{high}-}(P,T)..i']\}$. By $\delta \in [0...\delta_{f,\ell}^{\operatorname{high}-}(P,T))$, we thus have $\operatorname{SA}[q], \operatorname{SA}[q'] \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$. Consequently, $\operatorname{SA}[q], \operatorname{SA}[q'] \in \mathsf{R}_{f,s,H}^{-}(\tau,T)$ and $\exp_f(\tau,T,\operatorname{SA}[q]) = \exp_f(\tau,T,\operatorname{SA}[q']) = k_2$. Since for every $t \in \mathsf{R}(\tau,T)$, it holds $e_f^{\operatorname{till}}(\tau,T,t) t = \operatorname{head}_f(\tau,T,t) + |\operatorname{root}_f(\tau,T,t)| \cdot \exp_f(\tau,T,t)$, we thus have $e_f^{\operatorname{tull}}(\tau,T,\operatorname{SA}[q]) \operatorname{SA}[q] = e_f^{\operatorname{tull}}(\tau,T,\operatorname{SA}[q']) \operatorname{SA}[q']$. By definition of SA , we have $\operatorname{SA}[q] \neq \operatorname{SA}[q']$. Assume without the loss of generality that $\operatorname{SA}[q] < \operatorname{SA}[q']$, and suppose that $[\operatorname{SA}[q]...\operatorname{SA}[q']] \subseteq \mathsf{R}(\tau,T)$. By Lemma 8.48, we then have $e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[q]) = e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[q'])$, which implies $\operatorname{SA}[q] = \operatorname{SA}[q']$, a contradiction. We thus cannot have $[\operatorname{SA}[q]...\operatorname{SA}[q']] \subseteq \mathsf{R}(\tau,T)$, i.e., there exists $t \in (\operatorname{SA}[q]...\operatorname{SA}[q'])$ satisfying $t \notin \mathsf{R}(\tau,T)$. By Lemma 8.49(2), we then obtain $b(\tau,T,\operatorname{SA}[q]) \leq \operatorname{SA}[q] < t < b(\tau,T,\operatorname{SA}[q'])$. In particular, $b(\tau,T,\operatorname{SA}[q]) \neq b(\tau,T,\operatorname{SA}[q'])$.
 - Second, we show that $B\subseteq A'\backslash A$. Let $i''\in [i'-\delta\dots i']$ and $t=b(\tau,T,\mathrm{SA}[i''])$. To show $t\in A'\backslash A$, we need to prove that $t\in R'_{f,H}(\tau,T), x\leq e^{\mathrm{full}}_{\mathrm{full}}(\tau,T,t)-t,$ and $T^{\infty}[e^{\mathrm{full}}_{\mathrm{full}}(\tau,T,t)\dots e^{\mathrm{full}}_{\mathrm{full}}(\tau,T,t)+7\tau)\succeq y_u$. Above, we observed that it holds $\mathrm{SA}[i'']\in\mathrm{Pos}^{\mathrm{high}-}_{f,\ell}(P,T)$. This implies $\mathrm{SA}[i'']\in\mathrm{R}^-_{f,s,H}(\tau,T)$. On the other hand, by Lemma 8.49(2), we have $[t\dots\mathrm{SA}[i'']]\subseteq\mathrm{R}(\tau,T)$. Thus, by Lemma 8.48, we obtain $t\in\mathrm{R}^-_{f,H}(\tau,T)$. Since Lemma 8.49(2) also implies $t\in\mathrm{R}'(\tau,T)$, we thus obtain $t\in\mathrm{R}'^-_{f,H}(\tau,T)$. Next, observe that $\mathrm{SA}[i'']\in\mathrm{Pos}^{\mathrm{high}-}_{f,\ell}(P,T)$ also implies $\exp_f(\tau,T,\mathrm{SA}[i''])=k_2$. Combining this with $\mathrm{SA}[i'']\in\mathrm{R}_{f,s,H}(\tau,T)$ yields $e^{\mathrm{full}}_{t}(\tau,T,\mathrm{SA}[i''])-\mathrm{SA}[i'']=\mathrm{head}_{f}(\tau,T,\mathrm{SA}[i''])+\exp_{f}(\tau,T,\mathrm{SA}[i''])|\mathrm{proot}_{f}(\tau,T,\mathrm{SA}[i''])|=s+k_2p=\mathrm{head}_{f}(\tau,P)+\exp_{f}^{\mathrm{tut}}(\tau,P,2\ell)|\mathrm{root}_{f}(\tau,P)|=e^{\mathrm{tut}}_{f}(\tau,P,2\ell)-1=x.$ Note now that we have $t\leq\mathrm{SA}[i'']$. On the other hand, by $[t\dots\mathrm{SA}[i'']]\subseteq\mathrm{R}(\tau,T)$ and Lemma 8.48, it holds $e^{\mathrm{full}}_{f}(\tau,T,t)=e^{\mathrm{full}}_{f}(\tau,T,\mathrm{SA}[i''])$. Thus, $x=e^{\mathrm{full}}_{f}(\tau,T,\mathrm{SA}[i''])-\mathrm{SA}[i'']=e^{\mathrm{full}}_{f}(\tau,T,t)-\mathrm{SA}[i'']\leq e^{\mathrm{full}}_{f}(\tau,T,t)-t.$ It remains to prove $T^{\infty}[e^{\mathrm{full}}_{f}(\tau,T,t)\dots e^{\mathrm{full}}_{f}(\tau,T,t)+\tau)\succeq y_u.$ Recall that by $\delta\in[0\dots\delta_{f,\ell}^{f}(-P,T))$, we have $\mathrm{SA}[i'-\delta]+\mathrm{Pos}_{f,\ell}^{f}(-P,T)$. On the other hand, we also have $\mathrm{SA}[i'']\in\mathrm{Pos}_{f,\ell}^{f}(-P,T)$. Consequently, $\mathrm{SA}[i'-\delta]+\mathrm{SA}[i'']\in\mathrm{R}_{f,s,k_2,H}(\tau,T)$. This implies $\mathrm{LCE}_T(\mathrm{SA}[i'-\delta],\mathrm{SA}[i''])\geq s+k_2p=x.$ Note that by the uniqueness of T[n] in T and $i'-\delta\leq i''$, for every $x'\geq 0$, it holds $T^{\infty}[\mathrm{SA}[i'-\delta]+\mathrm{SA}[i'-\delta]+x')\preceq T^{\infty}[\mathrm{SA}[i'']+x.\mathrm{SA}[i'']+x'$. With $\mathrm{LCE}_T(\mathrm{SA}[i'-\delta],\mathrm{SA}[i''])\geq x$, this is equivalent to $T^{\infty}[\mathrm{SA}[i'-\delta]+x...\mathrm{SA}[i'']+x+1$. $T^{\infty}[\mathrm{SA}[i'']+x+1$. In particular, for $x'=7\tau$, we thus have $y_u=T^{\infty}[e^{\mathrm{full}}_{f}(\tau,T,\mathrm{SA}[i''])\dots e^{\mathrm{full}}_{f}(\tau,T,\mathrm{SA}[i''])+\tau$.
 - (b) Next, we show $c-\delta \leq \text{weight-count}_{\overline{f}}(x,n,y_u)$. By Lemma 8.81(2), we have weight-count_{\overline{f}}(x,n,y_u) = |A|, where $A = \{t \in \mathsf{R}_{f,H}^{\prime-}(\tau,T) : x \leq e_f^{\mathrm{full}}(\tau,T,t) t \text{ and } T^{\infty}[e_f^{\mathrm{full}}(\tau,T,t) ... e_f^{\mathrm{full}}(\tau,T,t) + 7\tau) \leq y_u\}$ (the requirement $x \leq 7\tau$ is proved above). On the other hand, by Lemma 8.81(3), we have c = |A'|, where $A' = \{t \in \mathsf{R}_{f,H}^{\prime-}(\tau,T) : x \leq e_f^{\mathrm{full}}(\tau,T,t) t\}$. Our claim is thus equivalent to $|A'| |A| \leq \delta$. By $A \subseteq A'$

our goal is to prove that $|A' \setminus A| \leq \delta$. We prove this in two steps.

- Let $B = \{b(\tau,T,t'): t' \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \text{ and } T^{\infty}[e_f^{\operatorname{full}}(\tau,T,t') \dots e_f^{\operatorname{full}}(\tau,T,t') + 7\tau) \succ y_u\}$. In the first step, we show that $A' \setminus A \subseteq B$. Let $q \in A' \setminus A$, i.e., $q \in \mathsf{R}'_{f,H}(\tau,T)$, $e_f^{\operatorname{full}}(\tau,T,q) q \geq x = e_f^{\operatorname{cut}}(\tau,P,2\ell)-1$, and $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,q) \dots e_f^{\operatorname{full}}(\tau,T,q) + 7\tau) \succ y_u$. Recall that $\delta \in [0 \dots \delta_{f,\ell}^{\operatorname{high}-}(P,T))$ implies that $\operatorname{SA}[i'-\delta] \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$. By Lemma 8.93 we thus obtain $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i'-\delta]) \dots e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i'-\delta]) + 7\tau) \succeq P_{\operatorname{sub}}$, where $P_{\operatorname{sub}} = P[e_f^{\operatorname{cut}}(\tau,P,2\ell) \dots \min(m,2\ell)]$. Consequently, $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,q) \dots e_f^{\operatorname{full}}(\tau,T,q) + 7\tau) \succeq y_u = T^{\infty}[e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i'-\delta]) \dots e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i'-\delta]) + 7\tau) \succeq P_{\operatorname{sub}}$. It therefore follows by Lemma 8.92 that, letting $t = e(\tau,T,q) q 3\tau + 2$, it holds $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \cap [q \dots q+t) = \{q'\}$, where $q' = e_f^{\operatorname{full}}(\tau,T,q) (e_f^{\operatorname{cut}}(\tau,P,2\ell)-1)$. Observe that $e_f^{\operatorname{full}}(\tau,T,q) q \geq e_f^{\operatorname{cut}}(\tau,P,2\ell) 1$ implies that $q' \geq q$. Combining this with $q' < q + t = e(\tau,T,q) 3\tau + 2$ and Lemma 8.49(1), implies that $[q \dots q'] \subseteq \operatorname{R}(\tau,T)$. Thus, by $q \in \operatorname{R}'(\tau,T)$ and Lemma 8.49(2), we have $b(\tau,T,q') = q$. Lastly, note that $[q \dots q'] \subseteq \operatorname{R}(\tau,T)$ and Lemma 8.48 imply $e_f^{\operatorname{full}}(\tau,T,q') = e_f^{\operatorname{full}}(\tau,T,q)$. Thus, $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,q') \dots e_f^{\operatorname{full}}(\tau,T,q') + 7\tau) = T^{\infty}[e_f^{\operatorname{full}}(\tau,T,q) \dots e_f^{\operatorname{full}}(\tau,T,q) + 7\tau) \succ y_u$. Hence, we obtain $q \in B$.
- Let us now denote $B'=\{b(\tau,T,\operatorname{SA}[i'']): i''\in (i'-\delta\ldots i']\}$. In the second step we show that $B\subseteq B'$. Consider any $i_1,i_2\in (i'-\delta_{f,\ell}^{\operatorname{high}-}(P,T)\ldots i']$ and assume that it holds $i_1< i_2$. As noted above, we then have $\operatorname{SA}[i_1],\operatorname{SA}[i_2]\in\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$. Consequently, $\operatorname{SA}[i_1],\operatorname{SA}[i_2]\in\operatorname{R}_{f,s,k_2,H}^-(\tau,T)$, and thus $e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i_1])-\operatorname{SA}[i_1]=e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i_2])-\operatorname{SA}[i_2]=s+k_2p=x$ and $T[\operatorname{SA}[i_1]\ldots e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i_1]))=T[\operatorname{SA}[i_2]\ldots e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i_2]))$. By definition of the suffix array and the uniqueness of T[n] in T, this implies $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i_1])\ldots e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i_1])+7\tau)\prec T[e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i_2])\ldots e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i_2])+7\tau)$. Consequently, for every $i''\in (i'-\delta_{f,\ell}^{\operatorname{full}}(\tau,T,\operatorname{SA}[i'-\delta))-e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i''])+7\tau) \succ y_u=T^{\infty}[e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i'-\delta])\ldots e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i'-\delta])+7\tau)$ implies $i''>i'-\delta$. Let us now consider any $t\in B$. Then, by definition, there exists $t'\in\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$ satisfying $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,t')\ldots e_f^{\operatorname{full}}(\tau,T,t')+7\tau)\succ y_u$ and $b(\tau,T,t')=t$. The assumption $t'\in\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$ implies that there exists $i''\in (i'-\delta_{f,\ell}^{\operatorname{high}-}(P,T)\ldots i']$ such that $t'=\operatorname{SA}[i'']$. By the above, the assumption $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i''])\ldots e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i''])+7\tau)\succ y_u$ implies that $i''\in (i'-\delta \ldots i']$. We thus have $b(\tau,T,\operatorname{SA}[i''])=b(\tau,T,t')=t\in B'$, which concludes the proof of $B\subseteq B'$.
- (c) Observe now that $|B'| \leq \delta$. By the above, we thus have $|A' \setminus A| \leq |B| \leq |B'| \leq \delta$.

We have thus proved that it holds $c-\delta\in (\text{weight-count}_{\mathcal{D}}(x,n,y_u)\dots \text{weight-count}_{\overline{\mathcal{D}}}^{\prec}(x,n,y_u)]$. By definition, this implies that weight-select $_{\mathcal{D}}(x,n,c-\delta)=\{j\in\mathbb{Z}:(x',y',w',j)\in\mathcal{P},x\leq x'< n,\text{ and }y'=y_u\}$. Let us thus consider some $j\in \text{weight-select}_{\mathcal{D}}(x,n,c-\delta)$. By the above, there exists $(x',y',w',j)\in\mathcal{P}$ satisfying $x\leq x'< n$ and $y'=y_u$. By Definition 6.5, the position j satisfies $j\in[1\dots n]$ and $T^{\infty}[j\dots j+7\tau)=y'=y_u=T^{\infty}[e_f^{\mathrm{full}}(\tau,T,\mathrm{SA}[i'-\delta])\dots e_f^{\mathrm{full}}(\tau,T,\mathrm{SA}[i'-\delta])+7\tau)=T^{\infty}[\mathrm{SA}[i'-\delta]+x\dots\mathrm{SA}[i'-\delta]+x+7\tau)$. Consequently, $T^{\infty}[j\dots j+7\tau-x)=T^{\infty}[\mathrm{SA}[i'-\delta]+x\dots\mathrm{SA}[i'-\delta]+\tau$. Next, we prove that $j-x\in[1\dots n]$ and $T[j-x\dots j)=T[\mathrm{SA}[i'-\delta]\dots\mathrm{SA}[i'-\delta]+x$. We proceed in three steps:

- (a) By definition of \$\mathcal{P}\$ = IntStrPoints_{7τ}(Seed⁻_{f,H}(τ,T),T) (see Definition 6.5), there exists (q, h) ∈ Seed⁻_{f,H}(τ,T) such that x' = h and T[∞][q · q + 7τ) = y_u. Furthermore, by Definition 6.5, position j satisfies T[∞][j 7τ · . j + 7τ) = T[∞][q 7τ · . q + 7τ). Recall also (see Section 8.4.4), that (q, h) ∈ Seed⁻_{f,H}(τ,T) in turn implies that there exists q' ∈ comp_{14τ}(R⁻_{f,H}(τ,T),T) ⊆ R⁻_{f,H}(τ,T) such that q = e^{full}_f(τ,T,q') and h = min(e^{full}_f(τ,T,q') b(τ,T,q'),7τ). Let us denote q" = q x.
 (b) Next, we prove that j x ∈ [1 · . n] and T[∞][j x · . j) = H'H^{k2}, where H' is a length-s suffix of H.
- (b) Next, we prove that $j-x\in[1..n]$ and $T^{\infty}[j-x..j)=H'H^{k_2}$, where H' is a length-s suffix of H. Above, we observed that x' satisfies $x\leq x'< n$. By $x'=h=\min(e_f^{\mathrm{full}}(\tau,T,q')-b(\tau,T,q'),7\tau)$, this implies $x\leq e_f^{\mathrm{full}}(\tau,T,q')-b(\tau,T,q')$, or equivalently, $q''=q-x=e_f^{\mathrm{full}}(\tau,T,q')-x\geq b(\tau,T,q')$. On the other hand, by $q'\in\mathsf{R}_{f,H}(\tau,T)$, Lemma 8.48, and Lemma 8.49(2), we have $b(\tau,T,q')\in\mathsf{R}_{f,H}(\tau,T)$. Thus, $b(\tau,T,q')\leq q''\leq e_f^{\mathrm{full}}(\tau,T,q')$ and $e_f^{\mathrm{full}}(\tau,T,q')-q''=x=s+k_2p$ imply that $T[q''..q)=T[q''..e_f^{\mathrm{full}}(\tau,T,q'))=H'H^{k_2}$. By $x\leq 2\ell\leq 7\tau$ and $T^{\infty}[q-7\tau..q)=T^{\infty}[j-7\tau..j)$, we thus obtain $T^{\infty}[j-x..j)=H'H^{k_2}$. Since H does not contain T[n], we thus obtain that $T^{\infty}[j-x..j)$ also does not contain T[n]. Combining this with $j\in[1..n]$, we thus obtain $j-x\in[1..n]$.
- (c) Lastly, observe that $e_f^{\text{full}}(\tau, T, \text{SA}[i' \delta]) \text{SA}[i' \delta] = x$ and $\text{SA}[i' \delta] \in \mathsf{R}_{f,s,H}(\tau,T)$ implies that

 $T[SA[i'-\delta]...SA[i'-\delta]+x) = T[SA[i'-\delta]...e_{f}^{ill}(\tau,T,SA[i'-\delta])) = H'H^{k_2}.$ By $T^{\infty}[j-x...j) = H'H^{k_2}$, we thus obtain $T^{\infty}[j-x..j) = T^{\infty}[SA[i'-\delta]..SA[i'-\delta] + x)$.

We have thus proved that $j - x \in [1 ... n]$ and $T^{\infty}[j - x ... j) = T^{\infty}[SA[i' - \delta] ... SA[i' - \delta] + x)$. Combining this with $T^{\infty}[j...j-x+7\tau)=T^{\infty}[\mathrm{SA}[i'-\delta]+x...\mathrm{SA}[i'-\delta]+7\tau)$, we thus obtain $T^{\infty}[j-x...j-x+7\tau)=$ $T^{\infty}[SA[i'-\delta]..SA[i'-\delta]+7\tau).$

3. Applying the claim from the second step for $\delta = i' - i$ (note that we have $\delta \in [0...\delta_{f,\ell}^{\text{high}-}(P,T))$, since in the first step we observed that $SA[i] \in Pos_{f,\ell}^{\text{high}-}(P,T)$ implies $i' - \delta_{f,\ell}^{\text{high}-}(P,T) < i \leq i'$), we obtain that any $j \in \text{weight-select}_{\mathcal{P}}(x, n, c - (i' - i))$ satisfies $j - x \in [1 ... n]$ and $T^{\infty}[j - x ... j - x + 7\tau) = 0$ $T^{\infty}[\operatorname{SA}[i'-\delta]...\operatorname{SA}[i'-\delta]+7\tau) = T^{\infty}[\operatorname{SA}[i]...\operatorname{SA}[i]+7\tau). \text{ By SA}[i] \in \operatorname{Occ}_{2\ell}(P,T), \ 2\ell \leq 7\tau, \text{ and Lemma 8.10,}$ we thus have $j-x \in \operatorname{Occ}_{2\ell}(P,T).$ On the other hand, $\operatorname{SA}[i] \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$ and Lemma 8.79 imply that $j-x \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T).$ We thus obtain $j-x \in \operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T).$

Lemma 8.122. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{type}(\tau, P) = -1$, $e(\tau, P) - 1 < \min(m, 2\ell)$, and T[n] does not occur in P[1 ... m). Denote $H = \operatorname{root}_f(\tau, P)$, $\mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f, H}^-(\tau, T), T)$, $i' = \operatorname{RangeBeg}_\ell(P, T) + \delta_{f, \ell}^{\operatorname{low}^-}(P, T) + \delta_{f, \ell}^{\operatorname{mid}^-}(P, T)$, $x = e_f^{\operatorname{full}}(\tau, P) - 1$, and $c = \operatorname{weight-count}_{\mathcal{P}}(x, n)$. Let $i \in [1 \dots n]$ be such that $\operatorname{SA}[i] \in \operatorname{Occ}_{2\ell}(P, T)$. Then:

- It holds $c (i' i) \in [1 ... c]$, i.e., weight-select_P(x, n, c (i' i)) is well-defined,
- Every position $j \in \text{weight-select}_{\mathcal{P}}(x, n, c (i' i))$ satisfies $j x \in \text{Occ}_{2\ell}(P, T)$.

 $\textit{Proof.} \text{ By Lemma 8.116, it holds } x = e_f^{\text{full}}(\tau, P) - 1 = e_f^{\text{cut}}(\tau, P, 2\ell) - 1 \text{ and } \text{SA}[i] \in \text{Pos}_{f,\ell}^{\text{high}-}(P, T). \text{ The lemma 8.116}$ claim thus follows by Lemma 8.121.

Lemma 8.123. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^+$ be a τ -periodic pattern satisfying $e(\tau, P) - 1 \geq 2\ell$. Denote $s = \text{head}_f(\tau, P)$, $H = \text{root}_f(\tau, P)$, p = |H|, and $k_2 = \lfloor \frac{2\ell - s}{p} \rfloor$. Then:

- 1. It holds $\mathsf{R}_{f,s,k_2+1,H}(\tau,T) \cup \mathsf{Pos}^{\mathsf{high}}_{f,\ell}(P,T) \subseteq \mathsf{Occ}_{2\ell}(P,T)$. 2. If $\mathsf{Occ}_{2\ell}(P,T) \neq \emptyset$, then $\mathsf{R}_{f,s,k_2+1,H}(\tau,T) \cup \mathsf{Pos}^{\mathsf{high}}_{f,\ell}(P,T) \neq \emptyset$.

Proof. Denote m=|P| and $t=\mathrm{tail}_f(\tau,P)$. Observe that the assumption $e(\tau,P)-1\geq 2\ell$ implies that $\exp_f^{\mathrm{cut}}(\tau,P,2\ell)=\min(\exp_f(\tau,P),\lfloor\frac{2\ell-s}{p}\rfloor)=\min(\lfloor\frac{e(\tau,P)-1-s}{p}\rfloor,\lfloor\frac{2\ell-s}{p}\rfloor)=\lfloor\frac{2\ell-s}{p}\rfloor=k_2$ and $m\geq e(\tau,P)-1\geq 2\ell$. Let us thus denote $P':=P[1\mathinner{.\,.} 2\ell]$. The assumption $e(\tau,P)-1\geq 2\ell$ then implies that, letting H'(resp. H'') be a length-s suffix (resp. length-t prefix) of H, we have $P' = H'H^{k_2}H''$. Consequently, by definition of $\operatorname{Occ}_{2\ell}(P,T)$, $j \in \operatorname{Occ}_{2\ell}(P,T)$ holds if and only if P' is a prefix of T[j ... n], i.e., $j \in \operatorname{Occ}(P',T)$. Thus, $\operatorname{Occ}_{2\ell}(P,T) = \operatorname{Occ}(P',T)$. Note also that P' is τ -periodic, head $f(\tau,P') = s$, root $f(\tau,P') = H$, $e(\tau, P') - 1 = 2\ell$, and type $(\tau, P') = -1$.

- 1. We show each of the two inclusions separately:
- Let $j \in \mathsf{R}_{f,s,k_2+1,H}(\tau,T)$. By definition, this implies that $H'H^{k_2+1}$ is a prefix of $T[j\mathinner{\ldotp\ldotp} n]$. Thus,
- $j \in \operatorname{Occ}(P',T) = \operatorname{Occ}_{2\ell}(P,T).$ Let $j \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T) = \operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T) \cup \operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T).$ Consider first the case $j \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T)$, i.e., $j \in \mathsf{R}_{f,s,k_2,H}^-(\tau,T)$ and either $T[j\mathinner{\ldotp\ldotp} n] \succeq P$ or $\operatorname{lcp}(P,T[j\mathinner{\ldotp\ldotp} n]) \geq 2\ell$. If the latter condition holds, i.e., $j \in \mathsf{R}_{f,s,k_2,H}^-(\tau,T)$ and either $T[j\mathinner{\ldotp\ldotp} n] \succeq P$ or $\operatorname{lcp}(P,T[j\mathinner{\ldotp\ldotp} n]) \geq 2\ell$. Note that $P \succeq P'$, then we immediately obtain $j \in \mathrm{Occ}_{2\ell}(P,T)$. Let us thus assume $T[j ... n] \succeq P$. Note that $P \succeq P'$, and hence $T[j ... n] \succeq P'$. We consider two cases. If T[j ... n] = P', then $j \in Occ(P', T) = Occ_{2\ell}(P, T)$. Let us thus assume $T[j ... n] \succ P'$. Observe that we then have $e(\tau, T, j) - j \ge e(\tau, P') - 1$, since otherwise, by Lemma 8.51(3), we would have $T[j ... n] \prec P'$, contradicting our assumption. Consequently, $e(\tau,T,j)-j\geq e(\tau,P')-1=2\ell.$ By $j\in\mathsf{R}_{f,s,H}(\tau,T)$, this implies that $H'H^{k_2}H''$ is a prefix of $T[j\mathinner{.\,.} n]$, i.e., $j\in\mathsf{Occ}(P',T)=\mathsf{Occ}_{2\ell}(P,T)$. The proof of the inclusion $\mathsf{Pos}^{\mathsf{high}+}_{f,\ell}(P,T)\subseteq\mathsf{Occ}_{2\ell}(P,T)$ is symmetric to the proof for $\mathsf{Pos}^{\mathsf{high}-}_{f,\ell}(P,T)$.
- 2. Let us now assume $\mathrm{Occ}_{2\ell}(P,T) \neq \emptyset$. Consider any $j \in \mathrm{Occ}_{2\ell}(P,T) = \mathrm{Occ}(P',T)$. This implies that $lcp(T[j..n], P') \ge 3\tau - 1$. Hence, by Lemma 8.50(1), we obtain $j \in R_{f,s,H}(\tau,T)$. Furthermore, observe that we must have $e(\tau, T, j) - j \ge e(\tau, P') - 1$, since otherwise, by Lemma 8.51(1), we would have lcp(P', T[j ... n]) =

 $\min(e(\tau,T,j)-j,e(\tau,P')-1)=e(\tau,T,j)-j< e(\tau,P')-1=2\ell, \text{ contradicting lcp}(P',T[j\mathinner{\ldotp\ldotp} n])=2\ell \text{ (implied by } j\in \operatorname{Occ}(P',T)). \text{ Consequently, } \exp_f(\tau,T,j)=\lfloor\frac{e(\tau,T,j)-j-s}{p}\rfloor\geq \lfloor\frac{e(\tau,P')-1-s}{p}\rfloor=\lfloor\frac{2\ell-s}{p}\rfloor=k_2. \text{ Let us thus consider two cases:}$

- First, assume that $\exp_f(\tau, T, j) = k_2$. Since we also have $\exp_f^{\text{cut}}(\tau, P, 2\ell) = k_2$, $j \in \mathsf{R}_{f,s,H}(\tau, T)$, and $\operatorname{lcp}(P, T[j \dots n]) \geq \operatorname{lcp}(P', T[j \dots n]) = 2\ell$, we immediately obtain that either $j \in \operatorname{Pos}_{f,\ell}^{\text{high}-}(P, T)$ or $j \in \operatorname{Pos}_{f,\ell}^{\text{high}+}(P, T)$. Thus, $\operatorname{Pos}_{f,\ell}^{\text{high}}(P, T) \neq \emptyset$.
- Let us now assume $\exp_f(\tau,T,j) > k_2$. Let $j' = e_f^{\mathrm{full}}(\tau,T,j) (k_2+1) \cdot p s$. We will prove that $j' \in \mathsf{R}_{f,s,k_2+1,H}(\tau,T)$. First, observe that $\exp_f(\tau,T,j) > k_2$ implies that $e_f^{\mathrm{full}}(\tau,T,j) j = \mathrm{head}_f(\tau,T,j) + \exp_f(\tau,T,j)|\mathrm{root}_f(\tau,T,j)| \geq s + (k_1+1) \cdot p$. In other words, $j' \geq j$. On the other hand, note that $k_2 = \lfloor \frac{2\ell-s}{p} \rfloor$ implies that $e(\tau,T,j) j' \geq e_f^{\mathrm{full}}(\tau,T,j) j' = s + (k_2+1) \cdot p = s + (\lfloor \frac{2\ell-s}{p} \rfloor + 1) \cdot p \geq s + \lceil \frac{2\ell-s}{p} \rceil \cdot p \geq s + (2\ell-s) = 2\ell \geq 3\tau 1$. We thus have $j' \in [j \cdot ..e(\tau,T,j) 3\tau 1]$. By Lemma 8.49(1), we therefore obtain $[j \cdot ..j'] \subseteq \mathsf{R}(\tau,T)$. This in turn, by Lemma 8.48, implies $\mathrm{root}_f(\tau,T,j') = \mathrm{root}_f(\tau,T,j) = H$ and $e_f^{\mathrm{full}}(\tau,T,j') = e_f^{\mathrm{full}}(\tau,T,j)$. Consequently, $\mathrm{head}_f(\tau,T,j') = (e_f^{\mathrm{full}}(\tau,T,j') j') \bmod p = (e_f^{\mathrm{full}}(\tau,T,j) j') \bmod p = (s + (k_2+1) \cdot p) \bmod p = s$ and $\mathrm{exp}_f(\tau,T,j') = \lfloor \frac{e_f^{\mathrm{full}}(\tau,T,j') j'}{p} \rfloor = \lfloor \frac{s + (k_2+1) \cdot p}{p} \rfloor = k_2 + 1$. We have thus proved that $j' \in \mathsf{R}_{f,s,k_2+1,H}(\tau,T)$. Thus, $\mathsf{R}_{f,s,k_2+1,H}(\tau,T) \neq \emptyset$.

Lemma 8.124. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that T[n] does not occur in P[1..m), $e(\tau,P)-1 \geq 2\ell$, and $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \neq \emptyset$. Denote $s = \operatorname{head}_f(\tau,P)$, $H = \operatorname{root}_f(\tau,P)$, p = |H|, $k_2 = \lfloor \frac{2\ell-s}{p} \rfloor$, $\mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f,H}^-(\tau,T),T)$, $x = s + k_2 p$, and $c = \operatorname{weight-count}_{\mathcal{P}}(x,n)$. Let $c' \in [0...\delta_{f,\ell}^{\operatorname{high}-}(P,T))$. Then:

- It holds $c-c' \in [1 ... c]$, i.e., weight-select_P(x, n, c-c') is well-defined.
- Every position $j \in \text{weight-select}_{\mathcal{P}}(x, n, c c')$ satisfies $j x \in \text{Pos}_{f \ell}^{\text{high-}}(P, T)$.

Proof. First, observe that $m \ge e(\tau,P)-1 \ge 2\ell$. Let us thus denote $P' := P[1..2\ell]$. It holds $\log(P',P) = 2\ell \ge 3\tau-1$. By Lemma 8.45, we thus obtain that P' is τ -periodic and it holds $\operatorname{head}_f(\tau,P') = \operatorname{head}_f(\tau,P) = s$ and $\operatorname{root}_f(\tau,P') = \operatorname{root}_f(\tau,P) = H$. Observe also that by definition of $e(\tau,P)$ and the assumption $e(\tau,P)-1 \ge 2\ell$, we have $e(\tau,P)-1 = p + \operatorname{lcp}(P[1..m], P[1+p..m]) \ge 2\ell$. Thus, $p + \operatorname{lcp}(P[1..2\ell], P[1+p..2\ell]) = 2\ell$, and hence $e(\tau,P')-1 = p + \operatorname{lcp}(P'[1..2\ell], P'[1+p..2\ell]) = p + \operatorname{lcp}(P[1..2\ell], P[1+p..2\ell]) = 2\ell$. Consequently, $\operatorname{type}(\tau,P') = -1$ and $\operatorname{exp}_f(\tau,P') = \frac{e(\tau,P')-1-\operatorname{head}_f(\tau,P')}{|\operatorname{root}_f(\tau,P')|} = \frac{2\ell-s}{p} = k_2$. Note also that then $\operatorname{exp}_f^{\mathrm{cut}}(\tau,P',2\ell) = \min(\operatorname{exp}_f(\tau,P'), \lfloor \frac{2\ell-s}{p} \rfloor) = k_2$, which in turn implies $e_f^{\mathrm{cut}}(\tau,P',2\ell)-1 = s + \operatorname{exp}_f^{\mathrm{cut}}(\tau,P',2\ell) = s + k_2p = x$. Next, observe that by Lemma 8.72(3), it holds $\delta_{f,\ell}^{\mathrm{high}-}(P',T) = \delta_{f,\ell}^{\mathrm{high}-}(P'[1..2\ell],T) = \delta_{f,\ell}^{\mathrm{high}-}(P,T)$. Denote $i' = \operatorname{RangeBeg}_\ell(P',T) + \delta_{f,\ell}^{\mathrm{high}-}(P',T) + \delta_{f,\ell}^{\mathrm{high}-}(P',T) = \delta_{f,\ell}$

Lemma 8.125. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that T[n] does not occur in P[1..m), $e(\tau,P)-1 \geq 2\ell$, and $\operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P,T) \neq \emptyset$. Denote $s = \operatorname{head}_f(\tau,P)$, $H = \operatorname{root}_f(\tau,P)$, p = |H|, $k_2 = \lfloor \frac{2\ell-s}{p} \rfloor$, $\mathcal{P} = \operatorname{IntStrPoints}_{\tau_\tau}(\operatorname{Seed}_{f,H}^-(\tau,T),T)$, $x = s + k_2 p$, and $c = \operatorname{weight-count}_{\mathcal{P}}(x,n)$. Then, $c \geq 1$ and every position $j \in \operatorname{weight-select}_{\mathcal{P}}(x,n,c)$ satisfies $j - x \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P,T)$.

Proof. The result follows by Lemma 8.124 with c' = 0.

Lemma 8.126. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $i \in [1..n]$ be such that $SA[i] \in R^-(\tau, T)$ and $e(\tau, T, SA[i]) - SA[i] < 2\ell$. Denote $H = root_f(\tau, T, SA[i])$, $\mathcal{P} = root_f(\tau, T, SA[i])$

 $\operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f,H}^{-}(\tau,T),T) \ (\operatorname{Definition} \ \textbf{6.5}), \ i' = \operatorname{RangeBeg}_{\ell}(\operatorname{SA}[i],T) + \delta_{f,\ell}^{\operatorname{low}-}(\operatorname{SA}[i],T) + \delta_{f,\ell}^{\operatorname{mid}-}(\operatorname{SA}[i],T),$ $x = e_f^{\text{full}}(\tau, T, SA[i]) - SA[i], \text{ and } c = \text{weight-count}_{\mathcal{P}}(x, n). \text{ Then:}$

- It holds $c (i' i) \in [1 ... c]$, i.e., weight-select_P(x, n, c (i' i)) is well-defined,
- Every position $j \in \text{weight-select}_{\mathcal{P}}(x, n, c (i' i))$ satisfies $j x \in \text{Occ}_{2\ell}(SA[i], T)$.

Proof. Denote P := T[SA[i] ... n] and m := |P| = n - SA[i] + 1. First, observe that $SA[i] \in R^-(\tau, T)$ implies that P is τ -periodic and type $(\tau, P) = -1$. Next, we prove that $e(\tau, P) - 1 < \min(m, 2\ell)$. By the uniqueness of T[n] in T, we have $e(\tau, T, SA[i]) \leq n$. Thus, $e(\tau, P) - 1 = e(\tau, T[SA[i] ... n]) - 1 =$ $e(\tau, T, SA[i]) - SA[i] \le n - SA[i] < m$. On the other hand, the assumption $e(\tau, T, SA[i]) - SA[i] < 2\ell$ implies $e(\tau, P) - 1 = e(\tau, T, SA[i]) - SA[i] < 2\ell$. We thus have $e(\tau, P) - 1 < \min(m, 2\ell)$. Next, note that by the uniqueness of T[n] in T, the symbol T[n] does not occur in P[1..m). Observe also that, by definition, we

- $$\begin{split} \bullet & \operatorname{RangeBeg}_{\ell}(P,T) = \operatorname{RangeBeg}_{\ell}(T[\operatorname{SA}[i] \mathinner{\ldotp\ldotp} n], T) = \operatorname{RangeBeg}_{\ell}(\operatorname{SA}[i], T), \\ \bullet & \delta_{f,\ell}^{\operatorname{low}^-}(P,T) = \delta_{f,\ell}^{\operatorname{low}^-}(T[\operatorname{SA}[i] \mathinner{\ldotp\ldotp} n], T) = \delta_{f,\ell}^{\operatorname{low}^-}(\operatorname{SA}[i], T), \\ \bullet & \delta_{f,\ell}^{\operatorname{mid}^-}(P,T) = \delta_{f,\ell}^{\operatorname{mid}^-}(T[\operatorname{SA}[i] \mathinner{\ldotp\ldotp} n], T) = \delta_{f,\ell}^{\operatorname{mid}^-}(\operatorname{SA}[i], T), \\ \bullet & \operatorname{root}_f(\tau,P) = \operatorname{root}_f(\tau,T[\operatorname{SA}[i] \mathinner{\ldotp\ldotp} n]) = \operatorname{root}_f(\tau,T,\operatorname{SA}[i]) = H, \text{ and} \\ \bullet & e_f^{\operatorname{full}}(\tau,P) 1 = e_f^{\operatorname{full}}(\tau,T[\operatorname{SA}[i] \mathinner{\ldotp\ldotp} n]) 1 = e_f^{\operatorname{full}}(\tau,T,\operatorname{SA}[i]) \operatorname{SA}[i] = x. \end{split}$$

Lastly, note that since $lcp(T[SA[i]..n], P) = m \ge min(m, 2\ell)$, it holds $SA[i] \in Occ_{2\ell}(P, T)$. The claim thus follows by Lemma 8.122 (recall that $\operatorname{Occ}_{2\ell}(P,T) = \operatorname{Occ}_{2\ell}(T[\operatorname{SA}[i] \dots n],T) = \operatorname{Occ}_{2\ell}(\operatorname{SA}[i],T)$).

Lemma 8.127. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be any necklace-consistent function. Let $j \in \mathbb{R}(\tau, T)$ be such that $e(\tau, T, j) - j \ge 2\ell$. Denote $s = \text{head}_f(\tau, T, j)$, $H = \text{root}_f(\tau, T, j)$, p = |H|, and $k_2 = \lfloor \frac{2\ell - s}{n} \rfloor$. Then, it

- 1. $\mathsf{R}_{f,s,k_2+1,H}(\tau,T) \cup \mathsf{Pos}^{\mathsf{high}}_{f,\ell}(j,T) \subseteq \mathsf{Occ}_{2\ell}(j,T),$ 2. $\mathsf{R}_{f,s,k_2+1,H}(\tau,T) \cup \mathsf{Pos}^{\mathsf{high}}_{f,\ell}(j,T) \neq \emptyset.$

Proof. Denote P:=T[j..n]. First, observe that, by definition, $j \in R(\tau,T)$ implies that P is τ -periodic and it holds $\operatorname{head}_f(\tau, P) = \operatorname{head}_f(\tau, T[j \dots n]) = \operatorname{head}_f(\tau, T, j) = s$ and $\operatorname{root}_f(\tau, P) = \operatorname{root}_f(\tau, T[j \dots n]) = s$ $\operatorname{root}_f(\tau,T,j)=H$. Next, note that $e(\tau,P)-1=e(\tau,T[j\mathinner{.\,.} n])-1=e(\tau,T,j)-j\geq 2\ell$. Lastly, note that by definition, for every $j' \in [1..n]$ and $k \geq 0$, it holds $j' \in \text{Occ}_k(T[j'..n], T)$. In particular, $j \in \text{Occ}_k(T[j'..n], T)$. $\operatorname{Occ}_{2\ell}(T[j..n],T) = \operatorname{Occ}_{2\ell}(P,T)$. Thus, $\operatorname{Occ}_{2\ell}(P,T) \neq \emptyset$. Combining the above, it follows by Lemma 8.123 that $\mathsf{R}_{f,s,k_2+1,H}(\tau,T) \cup \mathsf{Pos}^{\mathsf{high}}_{f,\ell}(P,T) \subseteq \mathsf{Occ}_{\ell}(P,T)$ and $\mathsf{R}_{f,s,k_2+1,H}(\tau,T) \cup \mathsf{Pos}^{\mathsf{high}}_{f,\ell}(P,T) \neq \emptyset$. To obtain the claim, it remains to observe that $\mathsf{Pos}^{\mathsf{high}}_{f,\ell}(P,T) = \mathsf{Pos}^{\mathsf{high}}_{f,\ell}(T[j\mathinner{\ldotp\ldotp} n],T) = \mathsf{Pos}^{\mathsf{high}}_{f,\ell}(j,T)$ and $\mathsf{Occ}_{\ell}(P,T) = \mathsf{Occ}^{\mathsf{high}}_{f,\ell}(T[j\mathinner{\ldotp\ldotp\ldotp} n],T) = \mathsf{Occ}^{\mathsf{high}}_{f,\ell}(T[j\mathinner{\ldotp\ldotp\ldotp} n],T) = \mathsf{Occ}^{\mathsf{high}}_{f,\ell}(T[j\mathinner{\ldotp\ldotp\ldotp\ldotp} n],T)$ $\operatorname{Occ}_{2\ell}(T[j ... n], T) = \operatorname{Occ}_{2\ell}(j, T).$

Query Algorithms

Proposition 8.128. Let $k \in [4 ... \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $i \in [1 ... n]$ be such that $SA[i] \in R^-(\tau,T)$ and $e(\tau,T,SA[i]) - SA[i] < 2\ell$. Given CompSAPeriodic(T), the value k, the position i, some $j \in \text{Occ}_{3\tau-1}(SA[i],T)$ satisfying $j = \min \text{Occ}_{2\ell}(j,T)$, and the values $\text{head}_f(\tau,T,SA[i])$, $|\operatorname{root}_f(\tau, T, \operatorname{SA}[i])|$, RangeBeg $_{\ell}(\operatorname{SA}[i], T)$, $\delta_{f,\ell}^{\operatorname{low}^-}(\operatorname{SA}[i], T)$, $\delta_{f,\ell}^{\operatorname{mid}^-}(\operatorname{SA}[i], T)$, and $\exp_f(\tau, T, \operatorname{SA}[i])$ as input, we can compute a position $j' \in \operatorname{Occ}_{2\ell}(\operatorname{SA}[i], T)$ in $\mathcal{O}(\log^{3+\epsilon} n)$ time.

Proof. Let us denote $s = \text{head}_f(\tau, T, SA[i]), H = \text{root}_f(\tau, T, SA[i]), p = |H|, k = \exp_f(\tau, T, SA[i]), b = \text{root}_f(\tau, T, SA[i])$ RangeBeg_{ℓ}(SA[i], T), $\delta_1 = \delta_{f,\ell}^{\text{low}-}$ (SA[i], T), $\delta_2 = \delta_{f,\ell}^{\text{mid}-}$ (SA[i], T), and $\mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f,H}^{-}(\tau,T),T)$. Using Proposition 8.90 and the position j as input, in $\mathcal{O}(\log n)$ time we retrieve the pointer to the structure from Proposition 6.6 for Seed $_{fH}^{\tau}(\tau,T)$ (note that $j \in \mathsf{R}_{f,H}(\tau,T)$ holds by Lemma 8.50(2)), i.e., performing weighted range queries on \mathcal{P} . Note that the pointer is not null, since we assumed $SA[i] \in R^-(\tau, T)$. Thus, $R_{f,H}^-(\tau, T) \neq \emptyset$, which implies $Seed_{f,H}^-(\tau, T) \neq \emptyset$. In $\mathcal{O}(1)$ time we now calculate $x := e_f^{\text{full}}(\tau, T, SA[i]) - SA[i] = s + kp$. Then, using Proposition 6.6, in $\mathcal{O}(\log^{2+\epsilon} n)$ time we compute $c = \text{weight-count}_{\mathcal{P}}(x, n)$. Next, in $\mathcal{O}(1)$ time we set $i' = \text{RangeBeg}_{\ell}(SA[i], T) + \delta_{f,\ell}^{\text{low}}(SA[i], T) + \delta_{f,\ell}^{\text{mid}}(SA[i], T) = b + \delta_1 + \delta_2$. Finally, using Proposition 6.6, in $\mathcal{O}(\log^{3+\epsilon} n)$ time we compute a position $j \in \mathsf{weight\text{-}select}_{\mathcal{P}}(x, n, c - (i' - i))$. By Lemma 8.126, it holds $j - x \in \mathsf{Occ}_{2\ell}(\mathsf{SA}[i], T)$. We thus return j' := j - x as the answer. In total, we spend $\mathcal{O}(\log^{3+\epsilon} n)$ time. \square

Proposition 8.129. Let $k \in [4 .. \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $j \in \mathsf{R}(\tau,T)$ be a position satisfying $j = \min \mathsf{Occ}_{2\ell}(j,T)$. Denote $s = \mathsf{head}_f(\tau,T,j)$ and $H = \mathsf{root}_f(\tau,T,j)$. Let H' be a length-s suffix of H and P be a length- 2ℓ prefix of $H'H^\infty$. Then, P is τ -periodic. Moreover, assuming $\mathsf{Occ}(P,T) \neq \emptyset$, given $\mathsf{CompSAPeriodic}(T)$, value k, the position j, and the values $\mathsf{head}_f(\tau,T,j)$, $|\mathsf{root}_f(\tau,T,j)|$, $\delta^{\mathsf{high}-}_{f,\ell}(P,T)$, and $\delta^{\mathsf{high}+}_{f,\ell}(P,T)$ as input, we can compute a position $j' \in \mathsf{Occ}(P,T)$ in $\mathcal{O}(\log^{3+\epsilon} n)$ time.

Proof. Denote $s = \text{head}_f(\tau, T, j)$ and $p = |\text{root}_f(\tau, T, j)|$. By Lemma 8.57, P is τ -periodic, T[n] does not occur in $P[1 \dots 2\ell)$, $e(\tau, P) - 1 = 2\ell$, $\text{head}_f(\tau, P) = s$, and $\text{root}_f(\tau, P) = H$ (hence in particular, $|\text{root}_f(\tau, P)| = p$). In $\mathcal{O}(1)$ time we compute the value $k_2 = \lfloor \frac{2\ell - s}{p} \rfloor$. We then initialize the set $\mathcal{C} := \emptyset$, and perform the following two steps:

- 1. Denote $\mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f,H}^-(\tau,T),T)$. First, using Proposition 8.90 and the position j as input, in $\mathcal{O}(\log n)$ time we retrieve the pointer μ_H^- to the structure from Proposition 6.6 for $\operatorname{Seed}_{f,H}^-(\tau,T)$, i.e., performing weighted range queries on \mathcal{P} . If μ_H^- is a null pointer, this step is finished (note that this implies $\mathsf{R}_{f,H}^-(\tau,T) = \emptyset$, and hence in particular $\operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P,T) = \mathsf{R}_{f,s,k_2+1,H}^-(\tau,T) = \emptyset$). Let us thus assume that μ_H^- is not null. We then perform the following two substeps:
 - First, we check if $\delta_{f,\ell}^{\text{high}-}(P,T) > 0$. If not, we finish this substep and move to the second substep. Otherwise, in $\mathcal{O}(1)$ time we compute the value $x = s + k_2 p$. Next, using Proposition 6.6 in $\mathcal{O}(\log^{2+\epsilon} n)$ time, we compute $c = \text{weight-count}_{\mathcal{P}}(x,n)$. Then, using again Proposition 6.6, in $\mathcal{O}(\log^{3+\epsilon} n)$ time we compute a position $q \in \text{weight-select}_{\mathcal{P}}(x,n,c)$. Finally, we add the position q x to the set \mathcal{C} . By Lemma 8.125, it holds $q x \in \text{Pos}_{f,\ell}^{\text{high}-}(P,T)$.
 - $\mathcal{O}(\log^{3+\epsilon}n)$ time we compute $c = \text{weight-count}_{\mathcal{P}}(x,n)$. Then, using again Proposition 6.6, in $\mathcal{O}(\log^{3+\epsilon}n)$ time we compute a position $q \in \text{weight-select}_{\mathcal{P}}(x,n,c)$. Finally, we add the position q-x to the set c. By Lemma 8.125, it holds $q-x \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P,T)$.

 In the second substep, we first let $k=k_2+1$ and $x=s+k_p$. Using Proposition 6.6 in $\mathcal{O}(\log^{2+\epsilon}n)$ time we compute $h=\operatorname{weight-count}_{\mathcal{P}}(x,n)$. Denote $k_{\min}=\left\lceil\frac{3\tau-1-s}{p}\right\rceil-1$ and $k_{\max}=\left\lfloor\frac{7\tau-s}{p}\right\rfloor$. Note that by $2\ell \geq \ell \geq 3\tau-1$ we then have $k=k_2+1=\left\lfloor\frac{2\ell-s}{p}\right\rfloor+1\geq \left\lceil\frac{2\ell-s}{p}\right\rceil \geq \left\lceil\frac{3\tau-1-s}{p}\right\rceil > k_{\min}$. On the other hand, $k=k_2+1=\left\lfloor\frac{2\ell-s+p}{p}\right\rfloor \leq \left\lfloor\frac{2\ell+|\tau/3|-s}{p}\right\rfloor \leq \left\lfloor\frac{7\tau-s}{p}\right\rfloor = k_{\max}$, where $2\ell+\lfloor \tau/3\rfloor \leq 7\tau$ follows by $\tau=\lfloor\frac{\ell}{3}\rfloor$ and $\ell\geq 16$. Consequently, by Lemma 8.82, it holds $h=\lfloor R_{f,s,k_2+1,H}^-(\tau,T) \rfloor$. If h=0, we finish this step. Let us thus assume h>0. By Lemma 8.84, we then have weight-count $_{\mathcal{P}}(x,n)\geq 1$. Using Proposition 6.6 in $\mathcal{O}(\log^{3+\epsilon}n)$ time we compute a position $q\in \operatorname{weight-select}_{\mathcal{P}}(x,n,1)$. We add q-x to the set c. By Lemma 8.84, it holds $q-x\in R_{f,s,k_2+1,H}^-(\tau,T)$.
- 2. In the second step, we perform the symmetric computation using the structure from Proposition 6.6 for $P_{f,H}^+(\tau,T)$ (see Section 8.4.4). This results in potentially inserting elements of $\operatorname{Pos}_{f,\ell}^{\operatorname{high+}}(P,T)$ and $\operatorname{R}_{f,s,k_2+1,H}^+(\tau,T)$ into \mathcal{C} .

Observe that during the above computation, for each of the four sets $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$, $\operatorname{Pos}_{f,\ell}^{\operatorname{high}+}(P,T)$, $\operatorname{R}_{f,s,k_2+1,H}^-(\tau,T)$, and $\operatorname{R}_{f,s,k_2+1,H}^+(\tau,T)$, we first check if the set is nonempty, and if so, we insert into $\mathcal C$ one of its elements. On the one hand, by Lemma 8.123, this implies that $\mathcal C \subseteq \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \cup \operatorname{Pos}_{f,\ell}^{\operatorname{high}+}(P,T) \cup \operatorname{R}_{f,s,k_2+1,H}^-(\tau,T) \subseteq \operatorname{Occ}_{\ell}(P,T) \cup \operatorname{R}_{f,s,k_2+1,H}^+(\tau,T) \subseteq \operatorname{Occ}_{\ell}(P,T) = \operatorname{Occ}_{\ell}(P,T)$. On the other hand, observe that the same lemma implies that $\operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T) \cup \operatorname{R}_{f,s,k_2+1,H}^-(\tau,T) \neq \emptyset$. Thus, by the observation about the above procedure, we have $\mathcal C \neq \emptyset$. Consequently, to finish the algorithm, we pick an arbitrary element from $\mathcal C$ and return as the answer. In total, we spend $\mathcal O(\log^{3+\epsilon} n)$ time.

Remark 8.130. Note that in Proposition 8.129, it is possible that $\mathcal{C} = \emptyset$ holds after the first step, even when the returned pointer μ_H^- is not null. This is because Lemma 8.123 guarantees only that $\operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T) \cup \operatorname{R}_{f,s,k_2+1,H}(\tau,T) \neq \emptyset$ (i.e., $\operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T) \cup \operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T) \cup \operatorname{R}_{f,s,k_2+1,H}(\tau,T) \cup \operatorname{R}_{f,s,k_2+1,H}(\tau,T) \neq \emptyset$). It is, however, possible that $\operatorname{Pos}_{f,\ell}^{\operatorname{high}}(P,T) \cup \operatorname{R}_{f,s,k_2+1,H}(\tau,T) = \emptyset$, even when $\operatorname{R}_{f,H}(\tau,T) \neq \emptyset$.

8.4.12 Computing a Position in a Cover

Combinatorial Properties

- **Lemma 8.131.** Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be a necklace-consistent function. Let $P \in \Sigma^m$ be a τ -periodic pattern such that $\operatorname{type}(\tau, P) = -1$ and T[n] does not occur in $P[1 \dots m)$. Assume $\operatorname{Occ}_{2\ell}(P, T) \cap \mathbb{C}$ $\operatorname{Pos}_{f,\ell}^{\operatorname{high-}}(P,T) \neq \emptyset. \ \ \operatorname{Denote} \ c = \max \Sigma, \ x_l = e_f^{\operatorname{cut}}(\tau,P,2\ell) - 1, \ y_l = P[e_f^{\operatorname{cut}}(\tau,P,2\ell) \dots \min(m,2\ell)], \ y_u = y_l c^{\infty},$ $H = \operatorname{root}_f(\tau, P), \text{ and } \mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f, H}^-(\tau, T), T). \text{ Then:}$

 - 1. It holds $\min \operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}^{\operatorname{high}^-}_{f,\ell}(P,T) = \operatorname{r-min}_{\mathcal{P}}(x_l,n,y_l,y_u) x_l$. 2. The position $j = \min \operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}^{\operatorname{high}^-}_{f,\ell}(P,T)$ satisfies $j = \min \operatorname{Occ}_{4\ell}(j,T)$.
- Proof. 1. By Lemma 8.115, it holds $\operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) = \{e_f^{\operatorname{full}}(\tau,T,j) x_l : j \in \mathsf{R}_{f,H}^{\prime-}(\tau,T), \ x_l \leq e_f^{\operatorname{full}}(\tau,T,j) j, \ \text{and} \ y_l \leq T^{\infty}[e_f^{\operatorname{full}}(\tau,T,j) \cdot e_f^{\operatorname{full}}(\tau,T,j) + 7\tau) \prec y_u\}.$ The same lemma also implies $|\operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)| = \operatorname{weight-count}_{\mathcal{P}}(x_l,n,y_l,y_u).$ Thus, $\operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \neq \emptyset$ implies weight-count $_{\mathcal{P}}(x_l,n,y_l,y_u) > 0$. Consequently, by Lemma 8.85, we have $\min \operatorname{Occ}_{2\ell}(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) = 0$ $\operatorname{\mathsf{r-min}}_{\mathcal{P}}(x_l,n,y_l,y_u)-x_l$. Note that Lemma 8.85 requires $x_l \in [0..7\tau]$, which holds here since $x_l =$ $e_f^{\text{cut}}(\tau, P, 2\ell) - 1 \le 2\ell$, and for $\tau = \lfloor \frac{\ell}{3} \rfloor$ and $\ell \ge 16$, it holds $2\ell \le 7\tau$.
- $2\ell + \tau \leq 3\ell$. Suppose now that $j \neq \min \operatorname{Occ}_{4\ell}(j,T)$. Then, there exists $j' \in [1...j)$ such that $j' \in \operatorname{Occ}_{4\ell}(j,T)$, or equivalently (by Lemma 8.11) $T^{\infty}[j ...j + 4\ell) = T^{\infty}[j' ...j' + 4\ell)$. By Lemma 8.55 (with $\delta = 0$), we thus obtain $j' \in \mathbb{R}^-_{f,s,k_2,H}(\tau,T)$. Observe now that by $j \in \operatorname{Occ}_2(P,T) \cap \operatorname{Pos}_{f,\ell}^{\operatorname{high}^-}(P,T)$ and Lemma 8.114, the pattern $P' := P[e_f^{\operatorname{cult}}(\tau,P,2\ell) ... \min(m,2\ell)]$ is a prefix of $T^{\infty}[e_f^{\operatorname{full}}(\tau,T,j) ... e_f^{\operatorname{full}}(\tau,T,j) + 7\tau)$. Note that $e_f^{\operatorname{full}}(\tau,T,j) + T^{\operatorname{full}}(\tau,T,j) = T^{\operatorname{full}}(\tau,T,j)$. $|P'|-j=s+k_2p+|P'|=e_f^{\rm cut}(\tau,P,2\ell)-1+(\min(m,2\ell)-e_f^{\rm cut}(\tau,P,2\ell)+1)=\min(m,2\ell)\leq 2\ell. \text{ This occurrence}$ of P' is therefore contained in $T^{\infty}[j...j+2\ell)$. Recall now that $e_f^{\rm full}(\tau,T,j')-j'=s+k_2p=e_f^{\rm full}(\tau,T,j)-j.$ Thus we obtain from $T^{\infty}[j...j+4\ell)=T^{\infty}[j'...j'+4\ell]$ that $T^{\infty}[e_f^{\rm full}(\tau,T,j')...e_f^{\rm full}(\tau,T,j')+|P'|)=P'.$ By $|P'|\leq 2\ell\leq 7\tau$ (where $2\ell\leq 7\tau$ follows by $\tau=\lfloor\frac{\ell}{3}\rfloor$ and $\ell\geq 16$), we thus obtain that P' is a prefix of $T^{\infty}[e_f^{\text{full}}(\tau, T, j') \dots e_f^{\text{full}}(\tau, T, j') + 7\tau)$. Combining this with $j' \in \mathsf{R}_{f,s,k_2,H}^{-}(\tau, T)$ implies by Lemma 8.114 that $j' \in \mathsf{Occ}_{2\ell}(P,T) \cap \mathsf{Pos}_{f,\ell}^{\text{high}-}(P,T)$, contradicting the definition of j.
- **Lemma 8.132.** Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be a necklace-consistent function. Let $P \in \Sigma^m$ be a au-periodic pattern such that $\operatorname{type}(\tau, P) = -1$, $e(\tau, P) - 1 < \min(m, 2\ell)$, and T[n] does not occur in P[1 ...m). Assume $\operatorname{Occ}_{2\ell}(P, T) \neq \emptyset$. Denote $c = \max \Sigma$, $x_l = e_f^{\operatorname{full}}(\tau, P) - 1$, $y_l = P[e_f^{\operatorname{full}}(\tau, P) ... \min(m, 2\ell)]$, $y_u = y_l c^{\infty}$, $H = \operatorname{root}_f(\tau, P)$, and $\mathcal{P} = \operatorname{IntStrPoints}_{7\tau}(\operatorname{Seed}_{f, H}^-(\tau, T), T)$. Then:
 - 1. It holds $\min \operatorname{Occ}_{2\ell}(P,T) = \operatorname{r-min}_{\mathcal{P}}(x_l, n, y_l, y_u) x_l$.
 - 2. The position $j = \min \operatorname{Occ}_{2\ell}(P,T)$ satisfies $j = \min \operatorname{Occ}_{4\ell}(j,T)$.
- Proof. 1. By Lemma 8.116, it holds $x_l = e_f^{\mathrm{full}}(\tau, P) 1 = e_f^{\mathrm{cut}}(\tau, P, 2\ell) 1$ and $\mathrm{Occ}_{2\ell}(P, T) \subseteq \mathrm{Pos}_{f,\ell}^{\mathrm{high}-}(P, T)$. Thus, we have $y_l = P[e_f^{\mathrm{full}}(\tau, P) \ldots \min(m, 2\ell)] = P[e_f^{\mathrm{cut}}(\tau, P, 2\ell) \ldots \min(m, 2\ell)]$. Consequently, it follows by Lemma 8.131(1) that $\min \mathrm{Occ}_{2\ell}(P, T) = \min \mathrm{Occ}_{2\ell}(P, T) \cap \mathrm{Pos}_{f,\ell}^{\mathrm{high}-}(P, T) = \mathrm{r-min}_{\mathcal{P}}(x_l, n, y_l, y_u) x_l$.

 2. By Lemma 8.116, it holds $\mathrm{Occ}_{2\ell}(P, T) \subseteq \mathrm{Pos}_{f,\ell}^{\mathrm{high}-}(P, T)$. Therefore, by Lemma 8.131(2), the position $j = \min \mathrm{Occ}_{2\ell}(P, T) = \min \mathrm{Occ}_{2\ell}(P, T) \cap \mathrm{Pos}_{f,\ell}^{\mathrm{high}-}(P, T)$ satisfies $j = \min \mathrm{Occ}_{4\ell}(j, T)$.
- **Lemma 8.133.** Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be a necklace-consistent function. Let $P \in \Sigma^m$ be a au-periodic pattern such that $\operatorname{type}(\tau,P)=-1$, $e(\tau,P)-1\geq 2\ell$, and T[n] does not occur in $P[1\ldots m)$. Assume $\operatorname{Pos}_{f,\ell}^{\operatorname{high-}}(P,T) \neq \emptyset$. Denote $c = \max \Sigma$, $x_l = e_f^{\operatorname{cut}}(\tau,P,2\ell) - 1$, $y_l = P[e_f^{\operatorname{cut}}(\tau,P,2\ell) \dots \min(m,2\ell)]$, and $y_u = y_l c^{\infty}$, $H = \text{root}_f(\tau, P)$, and $\mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f,H}^{-}(\tau, T), T)$. Then:

 - $\begin{array}{l} \text{1. It holds } \min \operatorname{Pos}^{\operatorname{high-}}_{f,\ell}(P,T) = \operatorname{r-min}_{\mathcal{P}}(x_l,n,y_l,y_u) x_l. \\ \text{2. The position } j = \min \operatorname{Pos}^{\operatorname{high-}}_{f,\ell}(P,T) \text{ satisfies } j = \min \operatorname{Occ}_{4\ell}(j,T). \end{array}$
- $\begin{array}{l} \textit{Proof.} \ 1. \ \ \text{By Lemma} \ 8.123, \ e(\tau,P)-1 \geq 2\ell \ \text{implies that} \ \operatorname{Pos}^{\operatorname{high-}}_{f,\ell}(P,T) \subseteq \operatorname{Occ}_{2\ell}(P,T). \ \ \text{Consequently,} \\ \operatorname{Pos}^{\operatorname{high-}}_{f,\ell}(P,T) \neq \emptyset \ \ \text{implies that} \ \operatorname{Pos}^{\operatorname{high-}}_{f,\ell}(P,T) = \operatorname{Pos}^{\operatorname{high-}}_{f,\ell}(P,T) \cap \operatorname{Occ}_{2\ell}(P,T) \neq \emptyset. \ \ \text{Thus, it follows by} \\ \operatorname{Lemma} \ 8.131(1), \ \operatorname{that} \ \min \operatorname{Pos}^{\operatorname{high-}}_{f,\ell}(P,T) = \min \operatorname{Pos}^{\operatorname{high-}}_{f,\ell}(P,T) \cap \operatorname{Occ}_{2\ell}(P,T) = \operatorname{r-min}_{\mathcal{P}}(x_l,n,y_l,y_u) x_l. \end{array}$

2. Combining $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \subseteq \operatorname{Occ}_{2\ell}(P,T)$, $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) = \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \cap \operatorname{Occ}_{2\ell}(P,T) \neq \emptyset$, it follows by Lemma 8.131(2), that the position $j = \min \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) = \min \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T) \cap \operatorname{Occ}_{2\ell}(P,T)$ satisfies $j = \min \operatorname{Occ}_{4\ell}(j, T).$

Lemma 8.134. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be a necklace-consistent function. Let $j \in \mathsf{R}^-(\tau,T)$ be such that $e(\tau, T, j) - j < 2\ell$. Denote $c = \max_{t} \Sigma$, $x_l = e_f^{\text{full}}(\tau, T, j) - j$, $y_l = T[e_f^{\text{full}}(\tau, T, j) \dots \min(n+1, j+2\ell))$, $y_u = y_l c^{\infty}$, $H = \text{root}_f(\tau, T, j)$, and $\mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f, H}^-(\tau, T), T)$. Then, we have $\text{Occ}_{2\ell}(j, T) \neq \emptyset$ and:

- 1. It holds $\min \operatorname{Occ}_{2\ell}(j,T) = \operatorname{r-min}_{\mathcal{P}}(x_l, n, y_l, y_u) x_l$.
- 2. The position $j = \min \operatorname{Occ}_{2\ell}(j,T)$ satisfies $j = \min \operatorname{Occ}_{4\ell}(j,T)$.

Proof. First, note that $j \in \operatorname{Occ}_{2\ell}(j,T) = \operatorname{Occ}_{2\ell}(T[j ... n],T) = \operatorname{Occ}_{2\ell}(P,T)$. Thus, $\operatorname{Occ}_{2\ell}(P,T) \neq \emptyset$.

- 1. Denote P := T[j ... n] and m := |P| = n j + 1. First, observe that $j \in \mathbb{R}^{-}(\tau, T)$ implies that P is τ -periodic and type $(\tau, P) = -1$. By the uniqueness of T[n] in T, $e(\tau, T, j) \leq n$. Thus, $e(\tau, P)$ $1 = e(\tau, T[j ... n]) - 1 = e(\tau, T, j) - j \le n - j < m$. On the other hand, $e(\tau, T, j) - j < 2\ell$ implies $e(\tau, P) - 1 = e(\tau, T, j) - j < 2\ell$. Consequently, $e(\tau, P) - 1 < \min(m, 2\ell)$. Next, note that by the uniqueness of T[n] in T, the symbol T[n] does not occur in P[1..m). Observe also that, by definition, we have:
 - $\operatorname{root}_f(\tau, P) = \operatorname{root}_f(\tau, T[j \dots n]) = \operatorname{root}_f(\tau, T, j) = H$, and

 - $e_f^{\text{full}}(\tau, P) 1 = e_f^{\text{full}}(\tau, T[j \dots n]) 1 = e_f^{\text{full}}(\tau, T, j) j = x_l,$ $P[e_f^{\text{full}}(\tau, P) \dots \min(m, 2\ell)] = T[j + e_f^{\text{full}}(\tau, P) 1 \dots j + \min(m, 2\ell) 1] = T[e_f^{\text{full}}(\tau, T, j) \dots (j 1) + \min(n (j 1), 2\ell)] = T[e_f^{\text{full}}(\tau, T, j) \dots \min(n, j 1 + 2\ell)] = T[e_f^{\text{full}}(\tau, T, j) \dots \min(n + 1, j + 2\ell)) = y_l.$

By the above, it follows by Lemma 8.132(1), that $\min \operatorname{Occ}_{2\ell}(P,T) = \operatorname{r-min}_{\mathcal{P}}(x_l,n,y_l,y_u) - x_l$. It remains to observe that $Occ_{2\ell}(P,T) = Occ_{2\ell}(T[j ... n], T) = Occ_{2\ell}(j,T)$.

2. Using the above properties of P, by Lemma 8.132(2) it follows that the position $j = \min \operatorname{Occ}_{2\ell}(P,T) =$ $\min \operatorname{Occ}_{2\ell}(j,T)$ satisfies $j = \min \operatorname{Occ}_{4\ell}(j,T)$.

Lemma 8.135. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be a necklace-consistent function. Let $j \in \mathsf{R}^-(\tau,T)$ be such that $e(\tau,T,j)-j \geq 2\ell$ and $j \in \mathsf{Pos}^{\mathsf{high}^-}_{f,\ell}(j,T)$. Denote $c = \max \Sigma$, $x_l = e_f^{\mathsf{full}}(\tau,T,j)-j$, $y_l = T[e_f^{\mathsf{full}}(\tau,T,j) \ldots \min(n+1,j+2\ell))$, $y_u = y_l c^{\infty}$, $H = \mathsf{root}_f(\tau,T,j)$, and $\mathcal{P} = \mathsf{IntStrPoints}_{7\tau}(\mathsf{Seed}^-_{f,H}(\tau,T),T)$. Then:

- 1. It holds $\min \operatorname{Pos}^{\operatorname{high-}}_{f,\ell}(j,T) = \operatorname{r-min}_{\mathcal{P}}(x_l,n,y_l,y_u) x_l$. 2. The position $j = \min \operatorname{Pos}^{\operatorname{high-}}_{f,\ell}(j,T)$ satisfies $j = \min \operatorname{Occ}_{4\ell}(j,T)$.

Proof. 1. Let P := T[j ... n] and m := |P| = n - j + 1. First, observe that $j \in \mathbb{R}^{-}(\tau, T)$ implies that Pis τ -periodic and type $(\tau, P) = -1$. By the uniqueness of T[n] in T, the symbol T[n] does not occur in P[1..m]. Observe also that, by definition, we have $\text{root}_f(\tau, P) = \text{root}_f(\tau, T[j..n]) = \text{root}_f(\tau, T, j) = H$. Next, observe also that, by definition, we have root $f(\tau,T) = \operatorname{root}_f(\tau,T) = \operatorname{root}_f($ $\min(n-(j-1),2\ell)] = T[e_f^{\mathrm{full}}(\tau,T,j) \dots \min(n,j-1+2\ell)] = T[e_f^{\mathrm{full}}(\tau,T,j) \dots \min(n+1,j+2\ell)] = y_l. \text{ Lastly, observe that } j \in \operatorname{Pos}_{f,\ell}^{\mathrm{high}}(j,T) \text{ implies that } \operatorname{Pos}_{f,\ell}^{\mathrm{high}}(P,T) = \operatorname{Pos}_{f,\ell}^{\mathrm{high}}(T[j\dots n],T) = \operatorname{Pos}_{f,\ell}^{\mathrm{high}}(j,T) \neq \emptyset.$ Putting all this together, by Lemma 8.133(1) we thus obtain min $\operatorname{Pos}_{f,\ell}^{\mathrm{high}}(j,T) = \min \operatorname{Pos}_{f,\ell}^{\mathrm{high}}(T[j\dots n],T) = \operatorname{Pos}_{f,\ell}^{\mathrm{high}}(T[j\dots n],T) = \operatorname{Pos}_{f,\ell}^{\mathrm{high}}(T[j\dots n],T) = \operatorname{Pos}_{f,\ell}^{\mathrm{high}}(T[j\dots n],T)$ $\min \operatorname{Pos}_{l,\ell}^{\operatorname{high}-}(P,T) = \operatorname{r-min}_{\mathcal{P}}(x_l, n, y_l, y_u) - x_l.$

2. By the above properties of P, combined with Lemma 8.133(2), we have that $j = \min \mathsf{Pos}^{\mathsf{high}-}_{f,\ell}(j,T) = \min \mathsf{Pos}^{\mathsf{high}-}_{f,\ell}(T[j\mathinner{\ldotp\ldotp} n],T) = \min \mathsf{Pos}^{\mathsf{high}-}_{f,\ell}(P,T)$ satisfies $j = \min \mathsf{Occ}_{4\ell}(j,T)$.

Lemma 8.136. Let $\ell \in [16..n)$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and f be a necklace-consistent function. Let $j \in \mathsf{R}^-(\tau,T)$ be such that $e(\tau,T,j)-j\geq 2\ell$. Denote $s=\operatorname{head}_f(\tau,T,j),\ H=\operatorname{root}_f(\tau,T,j),\ p=|H|,\ and\ k_2=\lfloor\frac{2\ell-s}{p}\rfloor$. If $\exp_f(\tau, T, j) = k_2$, then $j \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}}(j, T)$. Otherwise, it holds $\mathsf{R}_{f,s,k_2+1,H}^-(\tau, T) \neq \emptyset$.

Proof. First, observe that $e(\tau, T, j) - j \ge 2\ell$ and $s = \text{head}_f(\tau, T, j)$ imply that $\exp_f(\tau, T, j) = \lfloor \frac{e(\tau, T, j) - j - s}{n} \rfloor \ge 2\ell$ $\lfloor \frac{2\ell-s}{p} \rfloor = k_2. \text{ This implies that } \exp_f^{\text{cut}}(\tau, T, j, 2\ell) = \min(\exp_f(\tau, T, j), k_2) = k_2.$ Assume $\exp_f(\tau, T, j) = k_2 = \exp_f^{\text{cut}}(\tau, T, j, 2\ell).$ By Lemma 8.77, we then obtain $j \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(j, T).$

Let us now assume $\exp_f(\tau, T, j) \neq k_2$, i.e., denoting $k = \exp_f(\tau, T, j)$, we then have $k \geq k_2 + 1$. Let Let us now assume $\exp_f(\tau, I, j) \neq k_2$, i.e., denoting $k = \exp_f(\tau, T, j)$, we then have $k \geq k_2 + 1$. Let $j' := e_f^{\text{full}}(\tau, T, j) - s - (k_2 + 1)p$. We will prove that $j' \in \mathsf{R}_{f,s,k_2+1,H}^-(\tau,T)$. On the one hand, the assumption $k \geq k_2 + 1$ implies that $e_f^{\text{full}}(\tau, T, j) - j = s + kp \geq s + (k_2 + 1)p$. Thus, $j' = e_f^{\text{full}}(\tau, T, j) - s - (k_2 + 1)p \geq j$. On the other hand, we obtain from $k_2 = \lfloor \frac{2\ell - s}{p} \rfloor$ that $e(\tau, T, j) - j' \geq e_f^{\text{full}}(\tau, T, j) - j' = s + (k_2 + 1)p = s + (\lfloor \frac{2\ell - s}{p} \rfloor + 1)p \geq s + \lceil \frac{2\ell - s}{p} \rceil p \geq 2\ell \geq 3\tau - 1$. We thus have $j' \in [j ... e(\tau, T, j) - 3\tau + 1]$. By Lemma 8.49(1), this implies that $[j ... j'] \subseteq \mathsf{R}(\tau, T)$. Observe now that by Lemma 8.48, it holds $j' \in \mathsf{R}_{f,H}^-(\tau, T)$ and $e_f^{\text{full}}(\tau, T, j') = e_f^{\text{full}}(\tau, T, j)$. Thus, head $f(\tau, T, j') = \frac{e_f^{\text{full}}(\tau, T, j') - j'}{e_f^{\text{full}}(\tau, T, j') - j'} \mod p = (e_f^{\text{full}}(\tau, T, j) - j') \mod p = (s + (k_2 + 1)p) \mod p = s$ and $\exp_f(\tau, T, j') = \lfloor \frac{e_f^{\text{full}}(\tau, T, j') - j'}{p} \rfloor = \lfloor \frac{s + (k_2 + 1)p}{p} \rfloor = k_2 + 1$. We have thus proved that $j' \in \mathsf{R}_{f,s,k_2+1,H}^-(\tau, T)$. Hence, $\mathsf{R}_{f,s,k_2+1,H}^-(\tau, T) \neq \emptyset$. Hence, $\dot{\mathsf{R}}_{f,s,k_2+1,H}^-(\tau,T) \neq \emptyset$.

Query Algorithms

Proposition 8.137. Let $k \in [4 ... \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $j \in \mathbb{R}^-(\tau,T)$ be such that $e(\tau, T, j) - j < 2\ell$. Given CompSAPeriodic(T), the value k, the position j, some $j' \in Occ_{3\tau-1}(j, T)$ satisfying $j' = \min \operatorname{Occ}_{2\ell}(j',T)$, and the values $\operatorname{head}_f(\tau,T,j)$, $|\operatorname{root}_f(\tau,T,j)|$, and $e(\tau,T,j)$ as input, we can compute a position $j'' \in Occ_{2\ell}(j,T)$ satisfying $j'' = \min Occ_{4\ell}(j'',T)$ in $\mathcal{O}(\log^2 n)$ time.

Proof. Let us denote $s = \text{head}_f(\tau, T, j), H = \text{root}_f(\tau, T, j), p = |H|, k = \exp_f(\tau, T, j), \text{ and } \mathcal{P} = \text{root}_f(\tau, T, j)$ IntStrPoints_{7 τ}(Seed_{f,H}(τ , T), T) (Definition 6.5). Furthermore, let $c = \max \Sigma$, $x_l = e_f^{\text{full}}(\tau, T, j) - j$, $y_l = T[e_f^{\text{full}}(\tau, T, j) \dots \min(n+1, j+2\ell)), \text{ and } y_u = y_l c^{\infty}.$ Finally, let $j'' = \text{r-min}_{\mathcal{P}}(x_l, n, y_l, y_u).$ On the one hand, by Lemma 8.134(1), it holds $j'' = \min \operatorname{Occ}_{2\ell}(j,T)$. On the other hand, by Lemma 8.134(2), position j'' satisfies $j'' = \min \operatorname{Occ}_{4\ell}(j'', T)$. Thus, our goal is to compute $\operatorname{r-min}_{\mathcal{P}}(x_l, n, y_l, y_u)$. Using Proposition 8.90 and the position j' as input, in $\mathcal{O}(\log n)$ time we retrieve the pointer to the structure from Proposition 6.6 for $\operatorname{Seed}_{f,H}^{-}(\tau,T)$ (note that $j' \in R_{f,H}(\tau,T)$ holds by Lemma 8.50(2)), i.e., performing weighted range queries on \mathcal{P} . Note that the pointer is not null, since we assumed $j \in \mathbb{R}^-(\tau, T)$. Thus, it holds $\mathbb{R}^-_{f,H}(\tau, T) \neq \emptyset$, which in turn implies $\operatorname{Seed}_{f,H}^-(\tau, T) \neq \emptyset$. In $\mathcal{O}(1)$ time we now calculate $k := \exp_f(\tau, T, j) = \lfloor \frac{e(\tau, T, j) - s}{p} \rfloor$ and $x_l := e_f^{\operatorname{full}}(\tau, T, j) - j = \operatorname{head}_f(\tau, T, j) + \exp_f(\tau, T, j) \cdot |\operatorname{root}_f(\tau, T, j)| = s + kp$. Note that using CompSACore(T) (which is part of CompSAPeriodic(T)), we can lexicographically compare any two substrings of T^{∞} or $\overline{T^{\infty}}$ (specified with their starting positions and lengths) in $t_{\text{cmp}} = \mathcal{O}(\log n)$ time. In $\mathcal{O}(\log^{1+\epsilon} n + t_{\text{cmp}} \log n) = \mathcal{O}(\log^2 n)$ time, we thus compute and return $r\text{-min}_{\mathcal{P}}(x_l, n, y_l, y_u) - x_l$ using Proposition 6.6 with arguments $i = j + x_l$ and $q_r = \min(n+1, j+2\ell) - i$. In total, we spend $\mathcal{O}(\log^2 n)$

Proposition 8.138. Let $k \in [4 ... \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $j \in \mathbb{R}^-(\tau,T)$ be such that $e(\tau, T, j) - j \ge 2\ell$. Given CompSAPeriodic(T), the value k, the position j, some $j' \in \text{Occ}_{3\tau-1}(j, T)$ satisfying $j' = \min \operatorname{Occ}_{2\ell}(j',T)$, and the values $\operatorname{head}_f(\tau,T,j)$, $|\operatorname{root}_f(\tau,T,j)|$, and $e(\tau,T,j)$ as input, we can compute a position $j'' \in \text{Occ}_{2\ell}(j,T)$ satisfying $j'' = \min \text{Occ}_{4\ell}(j'',T)$ in $\mathcal{O}(\log^2 n)$ time.

 $Proof. \ \ \text{Denote} \ \ s = \text{head}_{f}(\tau, T, j), \ H = \text{root}_{f}(\tau, T, j), \ p = |H|, \ \mathcal{P} = \text{IntStrPoints}_{7\tau}(\text{Seed}_{f, H}^{-}(\tau, T), T), \ k = \text{IntSt$ $\exp_f(\tau, T, j)$, and $k_2 = \lfloor \frac{2\ell - s}{p} \rfloor$. We have two cases:

- 1. Let us first assume $k = k_2$. By Lemma 8.136, we then have $j \in \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(P,T)$. Consequently, letting $c = \max \Sigma$, $x_l = e_f^{\operatorname{full}}(\tau,T,j) j$, $y_l = T[e_f^{\operatorname{full}}(\tau,T,j) \dots \min(n+1,j+2\ell))$, $y_u = y_l c^{\infty}$, and $j'' = \operatorname{r-min}_{\mathcal{P}}(x_l,n,y_l,y_u) x_l$, by Lemma 8.135, the position j'' satisfies $j'' = \min \operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(j,T)$ and it holds $j'' = \min \operatorname{Occ}_{4\ell}(j'',T)$. Finally, note that by Lemma 8.127, we have $\operatorname{Pos}_{f,\ell}^{\operatorname{high}-}(j,T) \subseteq \operatorname{Occ}_{2\ell}(j,T)$. Thus, $j'' \in \mathrm{Occ}_{2\ell}(j,T)$.
- 2. Let us now assume $k \neq k_2$. By Lemma 8.136, we then have $\mathsf{R}^-_{f,s,k_2+1,H}(\tau,T) \neq \emptyset$. Then, letting $j'' = \mathsf{r\text{-}min}_{\mathcal{P}}(s+(k_2+1)p,n)-s-(k_2+1)p$, by Lemma 8.86, the position j'' satisfies $j'' = \min \mathsf{R}^-_{f,s,k_2+1,H}(\tau,T)$, and moreover, it holds $j'' = \min \mathsf{Occ}_{4\ell}(j'',T)$. Note that Lemma 8.86 requires that $\left\lceil \frac{3\tau-1-s}{\tau-1-s} \right\rceil 1 < k_2+1 \leq \left\lfloor \frac{7\tau-s}{\tau} \right\rfloor$. The first inequality holds since $3\tau-1 \leq 2\ell$ implies that $k_2+1 = \left\lfloor \frac{2\ell-s}{\tau-1-s} \right\rfloor + 1 \geq \left\lceil \frac{2\ell-s}{\tau-1-s} \right\rceil > \left\lceil \frac{3\tau-1-s}{\tau-1-s} \right\rceil 1$. For the second inequality, we note that $k_2+1 = \left\lfloor \frac{2\ell-s}{\tau-1-s} \right\rfloor + 1 = \left\lfloor \frac{2\ell-s}{\tau-1-s} \right\rfloor > \left\lceil \frac{3\tau-1-s}{\tau-1-s} \right\rceil 1$.

 $\lfloor \frac{2\ell-s+p}{p} \rfloor \leq \lfloor \frac{2\ell-s+\lfloor \tau/3 \rfloor}{p} \rfloor \leq \lfloor \frac{7\tau-s}{p} \rfloor, \text{ where } 2\ell + \lfloor \tau/3 \rfloor \leq 7\tau \text{ follows by } \tau = \lfloor \frac{\ell}{3} \rfloor \text{ and } \ell \geq 16. \text{ It remains to observe that by Lemma } 8.127, \text{ we have } \mathsf{R}_{f,s,k_2+1,H}^-(\tau,T) \subseteq \mathrm{Occ}_{2\ell}(j,T). \text{ Thus, } j'' \in \mathrm{Occ}_{2\ell}(j,T).$

The query algorithm thus proceeds as follows. By the uniqueness of T[n] in $T, j \in R(\tau, T)$ and $j' \in \operatorname{Occ}_{3\tau-1}(j,T)$ imply that $\operatorname{LCE}_T(j,j') \geq 3\tau-1$. By Lemma 8.50(2), we thus obtain $\operatorname{root}_f(\tau,T,j') = \operatorname{root}_f(\tau,T,j) = H$. Using Proposition 8.90 and the position j' as input, in $\mathcal{O}(\log n)$ time we retrieve the pointer to the structure from Proposition 6.6 for $\operatorname{Seed}_{f,H}^-(\tau,T)$, i.e., performing weighted range queries on \mathcal{P} . Note that the pointer is not null, since we assumed $j \in R^-(\tau,T)$. Thus, $R_{f,H}^-(\tau,T) \neq \emptyset$, which implies $\operatorname{Seed}_{f,H}^-(\tau,T) \neq \emptyset$. In $\mathcal{O}(1)$ time we now calculate $k := \exp_f(\tau,T,j) = \lfloor \frac{e(\tau,T,j)-j}{p} \rfloor$, $k_2 = \lfloor \frac{2\ell-s}{p} \rfloor$, and $x_l := e_f^{\operatorname{full}}(\tau,T,j) - j = s + kp$. Note that using $\operatorname{CompSACore}(T)$ (which is part of $\operatorname{CompSAPeriodic}(T)$), we can lexicographically compare any two substrings of T^∞ or T^∞ (specified with their starting positions and lengths) in $t_{\operatorname{cmp}} = \mathcal{O}(\log n)$ time. We consider two cases:

- 1. If $k = k_2$, then in $\mathcal{O}(\log^{1+\epsilon} n + t_{\text{cmp}} \log n) = \mathcal{O}(\log^2 n)$ time we compute and return $r\text{-min}_{\mathcal{P}}(x_l, n, y_l, y_u) x_l$ using Proposition 6.6 with arguments $i = j + x_l$ and $q_r = \min(n+1, j+2\ell) i$.
- 2. Otherwise, in $\mathcal{O}(\log^{1+\epsilon} n)$ time we compute and return $\operatorname{r-min}_{\mathcal{P}}(s+(k_2+1)p,n)-s-(k_2+1)p$ using Proposition 6.6.

In total, we spend $\mathcal{O}(\log^2 n)$ time.

Proposition 8.139. Let $k \in [4..\lceil \log n \rceil)$, $\ell = 2^k$, and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let $j \in R(\tau, T)$. Given CompSAPeriodic(T), the value k, the position j, and any $j' \in \operatorname{Occ}_{3\tau-1}(j,T)$ satisfying $j' = \min \operatorname{Occ}_{2\ell}(j',T)$ input, we can compute a position $j'' \in \operatorname{Occ}_{2\ell}(j,T)$ satisfying $j'' = \min \operatorname{Occ}_{4\ell}(j'',T)$ in $\mathcal{O}(\log^2 n)$ time.

Proof. Let $f = f_{\tau,T}$ (Definition 8.64). By the uniqueness of T[n] in $T, j \in R(\tau,T)$ and $j' \in Occ_{3\tau-1}(j,T)$ imply that $LCE_T(j,j') \geq 3\tau - 1$. By Lemma 8.50(2), we thus obtain $\operatorname{root}_f(\tau,T,j') = \operatorname{root}_f(\tau,T,j)$ and $\operatorname{head}_f(\tau,T,j') = \operatorname{head}_f(\tau,T,j)$. First, in $\mathcal{O}(\log n)$ time we compute $s = \operatorname{head}_f(\tau,T,j') = \operatorname{head}_f(\tau,T,j)$ and $p = |\operatorname{root}_f(\tau,T,j')| = |\operatorname{root}_f(\tau,T,j)|$ using Proposition 8.88. Next, in $\mathcal{O}(\log n)$ time we compute $t := e(\tau,T,j)$ using Proposition 8.89. Recall that using $\operatorname{CompSACore}(T)$ (which is part of $\operatorname{CompSAPeriodic}(T)$), we can access any symbol of T in $\mathcal{O}(\log n)$ time. In $\mathcal{O}(\log n)$ time we thus determine $\operatorname{type}(\tau,T,j)$ by comparing $T[e(\tau,T,j)]$ and $T[e(\tau,T,j)-p]$. Let us assume that $\operatorname{type}(\tau,T,j)=-1$ (the case of $\operatorname{type}(\tau,T,j)=+1$ is processed symmetrically). We then consider two cases. If $e(\tau,T,j)-j<2\ell$, we apply Proposition 8.137. Otherwise (i.e., $e(\tau,T,j)-j\geq2\ell$), we apply Proposition 8.138. In either case, we obtain a position $j'' \in \operatorname{Occ}_{2\ell}(j,T)$ satisfying $j'' = \min \operatorname{Occ}_{4\ell}(j'',T)$. In total, we spend $\mathcal{O}(\log^2 n)$ time.

8.4.13 Implementation of ISA Queries

Proposition 8.140. Let $k \in [4..\lceil \log n \rceil)$. Denote $\ell = 2^k$ and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let $j \in \mathsf{R}(\tau,T)$ be such that $j = \min \mathsf{Occ}_{2\ell}(j,T)$. Let $s = \mathsf{head}_f(\tau,T,j)$, $H = \mathsf{root}_f(\tau,T,j)$, H' be a length-s suffix of H, and P be a length- 2ℓ prefix of $H'H^{\infty}$. Given $\mathsf{CompSAPeriodic}(T)$ and the values k, j, s, |H|, $\mathsf{RangeBeg}_{\ell}(P,T)$, and $\mathsf{RangeEnd}_{\ell}(P,T)$ as input, in $\mathcal{O}(\log^{3+\epsilon}n)$ time we can compute $b = \mathsf{RangeBeg}(P,T)$ and $e = \mathsf{RangeBeg}(P,T)$. Moreover, if $b \neq e$ (i.e., $\mathsf{Occ}(P,T) \neq \emptyset$), we also return a position $j' \in \mathsf{Occ}(P,T)$ satisfying $j' = \min \mathsf{Occ}_{4\ell}(j',T)$.

Proof. By Lemma 8.57, it follows that P is τ -periodic, head $f(\tau, P) = s$, root $f(\tau, P) = H$, $e(\tau, P) - 1 = 2\ell$, and type $f(\tau, P) = -1$. Denote $f(\tau, P) = -1$. Denote $f(\tau, P) = -1$ and $f(\tau, P) = s$, root $f(\tau, P) = H$, $f(\tau, P) = H$, $f(\tau, P) = s$, root $f(\tau, P) = H$, $f(\tau, P) = s$, root $f(\tau,$

- 1. First, using Proposition 8.101, in $\mathcal{O}(\log^{2+\epsilon} n)$ time we compute $\delta_1^- := \delta_{f,\ell}^{\text{low}-}(P,T)$ and $\delta_3^- := \delta_{f,\ell}^{\text{high}-}(P,T)$. Using the symmetric variant of Proposition 8.101, in the same time we also compute $\delta_1^+ := \delta_{f,\ell}^{\text{high}+}(P,T)$ and $\delta_3^+ := \delta_{f,\ell}^{\text{high}+}(P,T)$.
- 2. Next, using Proposition 8.106 and its symmetric version, in $\mathcal{O}(\log^{2+\epsilon} n)$ time we compute $\delta_2^- := \delta_{f,\ell}^{\text{mid}-}(P,T)$ and $\delta_2^+ := \delta_{f,\ell}^{\text{mid}+}(P,T)$. By type $(\tau,P) = -1$ and Lemma 8.74, we then have $b = b' + \delta_1^- + \delta_2^- \delta_3^-$. By the symmetric version of Lemma 8.74, we also have $e = e' \delta_1^+ \delta_2^+ + \delta_3^+$. Thus, we

⁴Note that after inverting the lexicographic order, the type of P is still -1, because $e(\tau, P) = |P| + 1$. Thus, we can indeed use Lemma 8.74 without any changes (see also Remark 8.73).

- can determine b and e in $\mathcal{O}(1)$ time. If b = e, this completes the algorithm.
- 3. Let us thus assume $b \neq e$, i.e., $Occ(P,T) \neq \emptyset$. Using Proposition 8.129, in $\mathcal{O}(\log^{3+\epsilon} n)$ time we compute a position $j_2 \in \mathrm{Occ}(P,T)$.
- 4. Finally, using Proposition 8.139 and the position j_2 as input, in $\mathcal{O}(\log^2 n)$ time we compute a position $j' \in \mathrm{Occ}_{2\ell}(j_2, T)$ satisfying $j' = \min \mathrm{Occ}_{4\ell}(j', T)$. Note that since P does not contain symbol T[n], it holds $j_2 + 2\ell \le n$ and $T[j_2 ... j_2 + 2\ell) = P$. Thus, $Occ(P, T) = Occ_{2\ell}(j_2, T)$, and hence $j' \in Occ(P, T)$.

In total, we spend $\mathcal{O}(\log^{3+\epsilon} n)$ time.

Proposition 8.141. Let $k \in [4..\lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $j \in$ $\mathsf{R}^-(\tau,T)$ be such that $e(\tau,T,j)-j<2\ell$. Given $\mathsf{CompSAPeriodic}(T)$, the value k, the position j, the values RangeBeg_{\(\ell\)}(j,T), RangeEnd_{\(\ell\)}(j,T), head_{\(\ell\)} (τ,T,j) , $|\text{root}_f(\tau,T,j)|$, $e(\tau,T,j)$, and some position $j' \in$ $\operatorname{Occ}_{\ell}(j,T)$ satisfying $j' = \min \operatorname{Occ}_{2\ell}(j',T)$ as input, in $\mathcal{O}(\log^{2+\epsilon} n)$ time we can compute $\operatorname{RangeBeg}_{2\ell}(j,T)$, RangeEnd_{2ℓ}(j,T), and a position $j'' \in Occ_{2ℓ}(j,T)$ satisfying $j'' = min Occ_{4ℓ}(j'',T)$.

Proof. Denote $s = \operatorname{head}_f(\tau, T, j)$, $H = \operatorname{root}_f(\tau, T, j)$, p = |H|, $t = e(\tau, T, j)$, $b = \operatorname{RangeBeg}_\ell(j, T)$, and $e = \operatorname{RangeEnd}_\ell(j, T)$. We begin by computing $k := \exp_f(\tau, T, j) = \lfloor \frac{t-j-s}{p} \rfloor$, $k_1 := \exp_f^{\operatorname{cut}}(\tau, T, j, \ell) = \min(k, \lfloor \frac{\ell-s}{p} \rfloor)$, $k_2 := \exp_f^{\operatorname{cut}}(\tau, T, j, \ell) = \min(k, \lfloor \frac{2\ell-s}{p} \rfloor)$, and $q := e_f^{\operatorname{full}}(\tau, T, j) = j + s + kp$ in $\mathcal{O}(1)$ time. The rest of the query algorithm proceeds in four steps:

- First, using Proposition 8.100 (note that j ∈ Occ_ℓ(j, T) and j ∈ Occ_ℓ(j, T)), in O(log^{2+ϵ} n) time we compute the values δ₁ := δ_{f,ℓ}^{low-}(j, T) and δ₃ = δ_{f,ℓ}^{high-}(j, T).
 Next, using Proposition 8.105, in O(log^{2+ϵ} n) time we compute δ₂ := δ_{f,ℓ}^{mid-}(j, T). By Lemma 8.78, we can now in O(1) time determine the value b' := RangeBeg_{ℓℓ}(j, T) = RangeBeg_ℓ(j, T) + δ_{f,ℓ}^{low-}(j, T) + δ_{f,ℓ}^{low-}(j, T) + δ_{f,ℓ}^{low-}(j, T) $\delta_{f,\ell}^{\mathrm{mid}-}(j,T) - \delta_{f,\ell}^{\mathrm{high}-}(j,T) = b + \delta_1 + \delta_2 - \delta_3.$ 3. Next, using Proposition 8.120, in $\mathcal{O}(\log^{2+\epsilon} n)$ time we compute $t' := |\mathrm{Occ}_{2\ell}(j,T)|$. In $\mathcal{O}(1)$ time we
- then set $e' := \text{RangeEnd}_{2\ell}(j,T) = \text{RangeBeg}_{2\ell}(j,T) + |\text{Occ}_{2\ell}(j,T)| = e' + t'$.
- 4. Lastly, using Proposition 8.137, in $\mathcal{O}(\log^2 n)$ time we compute a position $j'' \in \operatorname{Occ}_{2\ell}(j,T)$ satisfying $j'' = \min \operatorname{Occ}_{4\ell}(j'', T).$

In total, we spend $\mathcal{O}(\log^{2+\epsilon} n)$ time.

Proposition 8.142. Let $k \in [4..\lceil \log n \rceil)$, $\ell = 2^k$, and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let $j \in \mathsf{R}(\tau, T)$. Given $\mathsf{CompSAPeriodic}(T)$, the value k, the position j, any $j' \in \text{Occ}_{\ell}(j,T)$ satisfying $j' = \min \text{Occ}_{2\ell}(j',T)$, and the values $\text{RangeBeg}_{\ell}(j,T)$ and Range $\operatorname{End}_{\ell}(j,T)$ as input, we can compute (Range $\operatorname{Beg}_{2\ell}(j,T)$, Range $\operatorname{End}_{2\ell}(j,T)$) and some position $j'' \in \operatorname{Occ}_{2\ell}(j,T)$ satisfying $j'' = \min \operatorname{Occ}_{4\ell}(j'',T)$ in $\mathcal{O}(\log^{3+\epsilon}n)$ time.

Proof. Let $f = f_{\tau,T}$ (Definition 8.64). By Lemma 8.50(2), it follows that $j' \in R(\tau,T)$, head $f(\tau,T,j') = f(\tau,T)$ $\operatorname{head}_f(\tau, T, j)$, and $\operatorname{root}_f(\tau, T, j') = \operatorname{root}_f(\tau, T, j)$. First, in $\mathcal{O}(\log n)$ time we compute $s := \operatorname{head}_f(\tau, T, j') = \operatorname{head}_f(\tau, T, j')$ $\operatorname{head}_f(\tau, T, j)$ and $p := |\operatorname{root}_f(\tau, T, j')| = |\operatorname{root}_f(\tau, T, j)|$ using Proposition 8.88. Next, in $\mathcal{O}(\log n)$ time we compute $t := e(\tau, T, j)$ using Proposition 8.89. We then consider two cases:

- Let us first assume $e(\tau,T,j)-j\geq 2\ell$. Let P be a length- 2ℓ prefix of $H'H^{\infty}$, Note that then $\operatorname{RangeBeg}_{\ell}(j,T) = \operatorname{RangeBeg}_{\ell}(P,T)$, and $\operatorname{RangeEnd}_{\ell}(j,T) = \operatorname{RangeEnd}_{\ell}(P,T)$. Using Proposition 8.140, in $\mathcal{O}(\log^{3+\epsilon} n)$ time we compute RangeBeg(P,T) = RangeBeg $_{2\ell}(j,T)$, RangeEnd(P,T) = RangeEnd_{2ℓ}(j, T), and some position $j'' \in \text{Occ}(P, T) = \text{Occ}_{2ℓ}(j, T)$ satisfying $j'' = \min \text{Occ}_{4ℓ}(j'', T)$.
- Let us now assume $e(\tau, T, j) j < 2\ell$. Recall that using CompSACore(T) (which is a component of CompSAPeriodic(T)), we can access any symbol of T in $\mathcal{O}(\log n)$ time. In $\mathcal{O}(\log n)$ time we thus compute type (τ, T, j) by comparing $T[e(\tau, T, j)] = T[t]$ with $T[e(\tau, T, j) - |\operatorname{root}_f(\tau, T, j)|] = T[t - p]$. Assume $\operatorname{type}(\tau, T, j) = -1$ (the case $\operatorname{type}(\tau, T, j) = +1$ is processed symmetrically). Using Proposition 8.141, in $\mathcal{O}(\log^{2+\epsilon} n)$ time we then compute $\operatorname{RangeBeg}_{2\ell}(j, T)$, Range $\operatorname{End}_{2\ell}(j, T)$, and some $j'' \in \operatorname{Occ}_{2\ell}(j, T)$ satisfying $j'' = \min \operatorname{Occ}_{4\ell}(j'', T)$.

In total, we spend $\mathcal{O}(\log^{3+\epsilon} n)$ time.

8.4.14 Implementation of SA Queries

Proposition 8.143. Let $k \in [4 ... \lceil \log n \rceil)$, $\ell = 2^k$, $\tau = \lfloor \frac{\ell}{3} \rfloor$, and $f = f_{\tau,T}$ (Definition 8.64). Let $i \in [1 ... n]$ be such that $SA[i] \in R^-(\tau,T)$ and $e(\tau,T,SA[i]) - SA[i] < 2\ell$. Given CompSAPeriodic(T), the value k, the position i, the values head $f(\tau, T, SA[i])$, $[root_f(\tau, T, SA[i])]$, $[RangeBeg_{\ell}(SA[i], T), RangeEnd_{\ell}(SA[i], T)]$ and some position $j \in \mathrm{Occ}_{\ell}(\mathrm{SA}[i],T)$ satisfying $j = \min \mathrm{Occ}_{2\ell}(j,T)$ as input, in $\mathcal{O}(\log^{3+\epsilon} n)$ time we $can\ compute\ \mathrm{RangeBeg}_{2\ell}(\mathrm{SA}[i],T),\ \mathrm{RangeEnd}_{2\ell}(\mathrm{SA}[i],T),\ and\ a\ position\ j'\in\mathrm{Occ}_{2\ell}(\mathrm{SA}[i],T)\ satisfying$ $j' = \min \operatorname{Occ}_{4\ell}(j', T).$

Proof. Let $b = \text{RangeBeg}_{\ell}(SA[i], T)$, $e = \text{RangeEnd}_{\ell}(SA[i], T)$, $s = \text{head}_{f}(\tau, T, SA[i])$, $H = \text{root}_{f}(\tau, T, SA[i])$, and p = |H|. Observe that by Lemma 8.50(2), it holds $j \in R(\tau, T)$, head $f(\tau, T, j) = \text{head } f(\tau, T, SA[i]) = s$, and $\operatorname{root}_f(\tau, T, j) = \operatorname{root}_f(\tau, T, \operatorname{SA}[i]) = H$. The query algorithm consists of eight steps:

- 1. Observe that $j \in \mathsf{R}_{f,s,H}(\tau,T)$ and $j \in \mathsf{Occ}_{\ell}(\mathsf{SA}[i],T)$ by Lemma 8.56(2) imply that $\mathsf{exp}_f^{\mathsf{cut}}(\tau,T,j,\ell) =$ $\exp_f^{\text{cut}}(\tau, T, \text{SA}[i], \ell)$. First, using Proposition 8.89 in $\mathcal{O}(\log n)$ time we compute $k_1 := \exp_f^{\text{cut}}(\tau, T, j, \ell) = 0$ $\exp_f^{\text{cut}}(\tau, T, \text{SA}[i], \ell).$
- 2. We apply Proposition 8.100 to position SA[i] and using the fact that $s = head_f(\tau, T, SA[i]), p =$ $|\operatorname{root}_f(\tau, T, \operatorname{SA}[i])|, j \in \operatorname{Occ}_\ell(\operatorname{SA}[i], T), \text{ and } k_1 = \exp_f^{\operatorname{cut}}(\tau, T, \operatorname{SA}[i], \ell), \text{ in } \mathcal{O}(\log^{2+\epsilon} n) \text{ time we compute}$ $\delta_1 := \delta_{f,\ell}^{\text{low}-}(SA[i], T).$
- 3. Using Proposition 8.112, in $\mathcal{O}(\log^{3+\epsilon} n)$ time, we compute $k := \exp_f(\tau, T, \mathrm{SA}[i])$. Observe that $e(\tau, T, \mathrm{SA}[i]) \mathrm{SA}[i] < 2\ell$ implies $\exp_f(\tau, T, \mathrm{SA}[i]) = \lfloor \frac{e(\tau, T, \mathrm{SA}[i]) \mathrm{SA}[i] s}{p} \rfloor \leq \lfloor \frac{2\ell s}{p} \rfloor$, and hence it holds $k_2 := \exp_f^{\text{cut}}(\tau, T, \text{SA}[i], 2\ell) = \min(\exp_f(\tau, T, \text{SA}[i]), \lfloor \frac{2\ell - s}{p} \rfloor) = \exp_f(\tau, T, \text{SA}[i]) = k.$
- 4. We apply Proposition 8.105 to SA[i], utilizing that $s = \text{head}_f(\tau, T, \text{SA}[i]), \ p = |\text{root}_f(\tau, T, \text{SA}[i])|, \ k_1 = \exp_f^{\text{cut}}(\tau, T, \text{SA}[i], \ell), \text{ and } k_2 = \exp_f^{\text{cut}}(\tau, T, \text{SA}[i], 2\ell), \text{ to compute } \delta_2 := \delta_{f,\ell}^{\text{mid}}(\text{SA}[i], T) \text{ in } \mathcal{O}(\log^{2+\epsilon} n)$
- 5. Using Proposition 8.128, in $\mathcal{O}(\log^{3+\epsilon} n)$ time we compute a position $j_2 \in \operatorname{Occ}_{2\ell}(\operatorname{SA}[i], T)$. 6. Using Proposition 8.100 for position $\operatorname{SA}[i]$, we compute $\delta_3 := \delta_{f,\ell}^{\operatorname{high}^-}(\operatorname{SA}[i], T)$ in $\mathcal{O}(\log^{2+\epsilon} n)$ time. Note that this time we utilize the fact that we have a position $j_2 \in \text{Occ}_{2\ell}(SA[i],T)$ and we know that $k_2 = \exp_f^{\text{cut}}(\tau, T, \text{SA}[i], 2\ell)$. After computing δ_3 , by Lemmas 8.7 and 8.78, we calculate $b' := \text{RangeBeg}_{2\ell}(\text{SA}[i], T) = \text{RangeBeg}_{\ell}(\text{SA}[i], T) + \delta_{f,\ell}^{\text{low}^-}(\text{SA}[i], T) + \delta_{f,\ell}^{\text{mid}^-}(\text{SA}[i], T) - \delta_{f,\ell}^{\text{high}^-}(\text{SA}[i], T) = 0$ $b+\delta_1+\delta_2-\delta_3$.
- 7. Observe that by Lemma 8.11, $j_2 \in \operatorname{Occ}_{2\ell}(\operatorname{SA}[i], T)$ implies $\operatorname{Occ}_{2\ell}(j_2, T) = \operatorname{Occ}_{2\ell}(\operatorname{SA}[i], T)$ (and hence, in particular, $|\operatorname{Occ}_{2\ell}(j_2,T)| = |\operatorname{Occ}_{2\ell}(\operatorname{SA}[i],T)|$. By Lemma 8.55, we then have $\operatorname{type}(\tau,T,j_2) =$ $type(\tau, T, SA[i]) = -1, \ e(\tau, T, j_2) - j_2 = e(\tau, T, SA[i]) - SA[i] < 2\ell, \ head_f(\tau, T, j_2) = head_f(\tau, T, SA[i])$ = s, $\operatorname{root}_f(\tau, T, j_2) = \operatorname{root}_f(\tau, T, \operatorname{SA}[i]) = H$ (in particular, $|\operatorname{root}_f(\tau, T, j_2)| = p$), and $\exp_f(\tau, T, j_2) = p$ $\exp_f(\tau, T, SA[i]) = k$. We then also have $j \in Occ_{3\tau-1}(j_2, T)$. Thus, using j_2, j, s, p , and k as input to Proposition 8.120, in $\mathcal{O}(\log^{2+\epsilon} n)$ time we compute $q := |\operatorname{Occ}_{2\ell}(j_2, T)| = |\operatorname{Occ}_{2\ell}(\operatorname{SA}[i], T)|$. In $\mathcal{O}(1)$ time we then calculate $e' := \text{RangeEnd}_{2\ell}(\text{SA}[i], T) = \text{RangeBeg}_{2\ell}(\text{SA}[i], T) + |\text{Occ}_{2\ell}(\text{SA}[i], T)| = b' + q$.
- 8. Using Proposition 8.139 for position j_2 and j, in $\mathcal{O}(\log^2 n)$ time we compute a position $j' \in \mathrm{Occ}_{2\ell}(j_2, T) =$ $\operatorname{Occ}_{2\ell}(\operatorname{SA}[i], T)$ satisfying $j' = \min \operatorname{Occ}_{4\ell}(j', T)$.

In total, we spend $\mathcal{O}(\log^{3+\epsilon} n)$ time.

Proposition 8.144. Let $k \in [4 .. \lceil \log n \rceil)$, $\ell = 2^k$, and $\tau = \lfloor \frac{\ell}{3} \rfloor$. Let $i \in [1 ... n]$ be such that $SA[i] \in R(\tau, T)$. Given CompSAPeriodic(T), the value k, the position i, the values RangeBeg_{ℓ}(SA[i], T), RangeEnd_{ℓ}(SA[i], T), and some position $j \in \mathrm{Occ}_{\ell}(\mathrm{SA}[i],T)$ satisfying $j = \min \mathrm{Occ}_{2\ell}(j,T)$ as input, in $\mathcal{O}(\log^{3+\epsilon} n)$ time we can compute RangeBeg_{2ℓ}(SA[i], T), RangeEnd_{2ℓ}(SA[i], T), and a position $j' \in Occ_{2ℓ}(SA[i], T)$ satisfying $j' = \min \operatorname{Occ}_{4\ell}(j', T).$

Proof. Let $f = f_{\tau,T}$ (Definition 8.64). Denote $b = \text{RangeBeg}_{\ell}(\text{SA}[i], T)$, and $e = \text{RangeEnd}_{\ell}(\text{SA}[i], T)$, and $H = \text{root}_f(\tau, T, SA[i])$. By Lemma 8.50(2), we have $j \in R(\tau, T)$, head $f(\tau, T, j) = \text{head}_f(\tau, T, SA[i])$ and $\operatorname{root}_f(\tau, T, j) = \operatorname{root}_f(\tau, T, \operatorname{SA}[i])$. First, in $\mathcal{O}(\log n)$ time we compute $s := \operatorname{head}_f(\tau, T, j) = \operatorname{head}_f(\tau, T, \operatorname{SA}[i])$ and $p := |\operatorname{root}_f(\tau, T, j)| = |\operatorname{root}_f(\tau, T, \operatorname{SA}[i])|$ using Proposition 8.88. Let $H' = H(p - s \cdot p)$ and P be a length- 2ℓ prefix of $H'H^{\infty}$. Using Proposition 8.140 and the position j as input, in $\mathcal{O}(\log^{3+\epsilon} n)$ time we compute $b_P := \text{RangeBeg}(P, T)$ and $e_P := \text{RangeEnd}(P, T)$. If $b_P \neq e_P$, we also obtain a position $j_P \in \text{Occ}(P, T)$ satisfying $j_P = \min \text{Occ}_{4\ell}(j_P, T)$. We then consider two cases:

- If $i \in (b_P ... e_P]$, then $SA[i] \in Occ(P,T)$. This implies $RangeBeg_{2\ell}(SA[i],T) = RangeBeg(P,T)$, $RangeEnd_{2\ell}(SA[i],T) = RangeEnd(P,T)$, and $Occ(P,T) = Occ_{2\ell}(SA[i],T)$. We thus return b_P , e_P , and j_P as the answer.
- Let us now assume $i \notin (b_P ... e_P]$. Moreover, let us assume $i \leq b_P$ (the case $i > e_P$ is handled symmetrically). This implies that $e(\tau, T, \mathrm{SA}[i]) \mathrm{SA}[i] < 2\ell$ and $\mathrm{SA}[i] \in \mathsf{R}^-(\tau, T)$. Consequently, we then compute $\mathrm{RangeBeg}_{2\ell}(\mathrm{SA}[i], T)$, $\mathrm{RangeEnd}_{2\ell}(\mathrm{SA}[i], T)$, and a position $j' \in \mathrm{Occ}_{2\ell}(\mathrm{SA}[i], T)$ satisfying $j' = \min \mathrm{Occ}_{4\ell}(j', T)$ using Proposition 8.143 in $\mathcal{O}(\log^{3+\epsilon} n)$ time.

In total, we spend $\mathcal{O}(\log^{3+\epsilon} n)$ time.

8.4.15 Construction Algorithm

Proposition 8.145. Given the LZ77 parsing of T, we can construct CompSAPeriodic(T) in $\mathcal{O}(\delta(T)\log^7 n)$ time.

Proof. First, using [KK20b, Theorem 6.11] (resp. [KK20b, Theorem 6.21]), in $\mathcal{O}(z(T)\log^4 n)$ time we construct a structure that, given any substring S of T (specified with its starting position and the length) in $\mathcal{O}(\log^3 n)$ time returns min $\operatorname{Occ}(S,T)$ (resp. $|\operatorname{Occ}(S,T)|$). Using [KK20b, Theorem 6.7], in $\mathcal{O}(z(T)\log^2 n)$ time we also construct a structure that, given any substring S of T (specified as above), in $\mathcal{O}(\log^3 n)$ time checks if $\operatorname{per}(S) \leq \frac{|S|}{2}$, and if so, returns $\operatorname{per}(S)$.

We then construct the components of the first part of CompSAPeriodic(T) (see Section 8.4.4) as follows:

- 1. In $\mathcal{O}(\delta(T)\log^7 n)$ time we construct CompSACore(T) using Proposition 8.18.
- 2. Next, we construct the arrays $A_{\text{root},i}$. Let $i \in [4..\lceil \log n \rceil)$. Denote $k = 14\tau_i$, $(p_j, t_j)_{j \in [1..n_{\text{runs},i}]} = \mathcal{I}(\text{comp}_k(\mathsf{R}(\tau_i, T), T))$, and

$$\begin{split} T'_{\text{comp}} &= \prod_{j=1,2,\dots,n_{\text{runs},i}} T[p_j \mathinner{\ldotp\ldotp} p_j + 2\tau_i - 1) T[n], \\ T''_{\text{comp}} &= \prod_{j=n_{\text{runs},i},\dots,2,1} T[p_j \mathinner{\ldotp\ldotp} p_j + 2\tau_i - 1) T[n], \\ Q &= \bigcup_{j \in \mathbb{R}(\tau_i,T)} \{ T[j+t\mathinner{\ldotp\ldotp} j + t + \text{per}(T[j\mathinner{\ldotp\ldotp} j + 3\tau_i - 1))) : t \in [0\mathinner{\ldotp\ldotp} \tau_i) \} \subseteq \Sigma^{\leq \tau_i/3}. \end{split}$$

Let also $g: \mathbb{Z} \to \mathbb{Z}$ be defined by $g(j) = 1 + 2\tau_i \lfloor \frac{j-1}{2\tau_i} \rfloor$. Consider any $j \in \mathsf{R}(\tau_i, T)$. Let $p = \mathsf{per}(T[j\mathinner{\ldotp\ldotp\ldotp} j + 3\tau_i - 1)), \ X = T[j\mathinner{\ldotp\ldotp\ldotp} j + p),$ and $s = \mathsf{min}\,\mathsf{Occ}(X^\infty[1\mathinner{\ldotp\ldotp\ldotp} \tau_i], T'_{\mathsf{comp}})$. Note that $X \in Q$, and hence by Definition 8.67, it holds $f_{k,\tau_i,T}(X) = T'_{\mathsf{comp}}[g(s)\mathinner{\ldotp\ldotp\ldotp} g(s) + |X|)$. Let us now denote $s' = \mathsf{max}\,\mathsf{Occ}(X^\infty[1\mathinner{\ldotp\ldotp\ldotp} \tau_i], T''_{\mathsf{comp}})$. Observe that due to length- $(2\tau_i - 1)$ substrings of T'_{comp} and T''_{comp} being separated by the sentinel symbol T[n], it holds $g(s') + g(s) = 2\tau_i(n_{\mathsf{runs},i} - 1) + 2$. This implies that $T'_{\mathsf{comp}}[g(s)\mathinner{\ldotp\ldotp\ldotp} g(s) + |X|) = T''_{\mathsf{comp}}[g(s')\mathinner{\ldotp\ldotp\ldotp} g(s') + |X|)$. Since we denoted $f_i = f_{\tau_i,T}$ (see Section 8.4.4), and since by $k = 14\tau_i \geq 3\tau_i - 1$ and Lemma 8.69, it holds $f_{k,\tau_i,T} = f_{\tau_i,T}$, we thus obtain

$$\operatorname{root}_{f_i}(\tau_i, T, j) = f_i(T[j ... j + p))$$

$$= f_i(X)$$

$$= f_{k, \tau_i, T}(X)$$

$$= T'_{\operatorname{comp}}[g(s) ... g(s) + p)$$

$$= T''_{\operatorname{comp}}[g(s') ... g(s') + p).$$

By the synchronization property of primitive strings [CHL07, Lemma 1.11], we therefore obtain head $f_i(\tau_i, T, j) = (g(s') - s') \mod p$. The key difficulty is thus computing s'. Observe that $T[j ... j + \tau_i) = f(j)$

⁵ Note that the rightmost occurrence of $X^{\infty}[1...\tau_i]$ in T''_{comp} may not correspond to the leftmost occurrence of $X^{\infty}[1...\tau_i]$ in T'_{comp} .

 $X^{\infty}[1..\tau_i]$. Thus, if we define $T_{\text{aux}} := T \cdot T''_{\text{comp}}$, then, letting $s'' = \max \text{Occ}_{\tau_i}(j, T_{\text{aux}})$, we have s'' = n + s'. Consequently, we have g(s') = g(s'' - n), and it holds

$$\operatorname{head}_{f_i}(\tau_i, T, j) = (g(s') - s') \bmod p$$
$$= (g(s'' - n) - (s'' - n)) \bmod p$$
$$= (1 + 2\tau_i \lfloor \frac{s'' - n - 1}{2\tau_i} \rfloor - (s'' - n)) \bmod p.$$

To compute $A_{\text{root},i}[1..n_{\text{runs},i}]$ we thus proceed as follows:

- (a) First, we construct an LZ77-like parsing of $T_{\rm aux}$. For this we first take the input LZ77 parsing of T. We then append the phrases encoding the remaining suffix, i.e., T''_{comp} . For $j = n_{\text{runs},i}, \ldots, 2, 1$ we encode the substring $T[p_j ... p_j + 2\tau_i - 1)T[n]$ using two phrases: the first is of length $2\tau_i - 1$ and has a source at position p_j (obtained from $A_{\text{runs},i}[j]$, which is a component of CompSACore(T); see Section 8.2.2). The second phrase is the single symbol. The resulting LZ77-like parsing has $z(T) + 2n_{\text{runs},i} = \mathcal{O}(z(T) + \delta(T)) = \mathcal{O}(z(T))$ phrases, where the subsequent inequalities follow by Lemma 8.14 and [KNP23]. Note also that $|T_{\text{aux}}| = |T| + |T''_{\text{comp}}| = n + 2\tau_i \cdot n_{\text{runs},i} = \mathcal{O}(n)$, where $\tau_i \cdot n_{\text{runs},i} = \mathcal{O}(n)$ follows by Lemma 8.14.
- (b) Using [KK20a, Theorem 6.12], in $\mathcal{O}(z(T)\log^4 n)$ time we then construct a data structure that lets us find the rightmost occurrence of every substring of $T_{\rm aux}$ (represented using its starting position and the length). More precisely, given any $x \in [1..|T_{\text{aux}}|]$ and t > 0, we can compute $\max \operatorname{Occ}_t(x, T_{\operatorname{aux}})$ in $\mathcal{O}(\log^3 n)$ time.
- (c) We now compute $A_{\text{root},i}[1...n_{\text{runs},i}]$. For $j=1,\ldots,n_{\text{runs},i}$ we perform the following steps:
 - i. First, from CompSACore(T), we obtain $(p_j, t_j) = A_{runs,i}[j]$.
 - ii. In $\mathcal{O}(\log^3 n)$ time we compute $p = \operatorname{per}(T[p_j \dots p_j + 3\tau_i 1))$ (note that $\operatorname{per}(T[p_j \dots p_j + 3\tau_i 1)) \leq 1$
 - iii. In $\mathcal{O}(\log^3 n)$ time we compute $s'' = \max \operatorname{Occ}_{\tau_i}(p_j, T_{\operatorname{aux}})$. Using the equation above, we then have head $f_i(\tau_i, T, p_j) = (1 + 2\tau_i \lfloor \frac{s'' n 1}{2\tau_i} \rfloor (s'' n)) \bmod p$.

Over all $j \in [1..n_{\text{runs},i}]$, we spend $\mathcal{O}(n_{\text{runs},i} \cdot \log^3 n) = \mathcal{O}(\delta(T) \log^3 n)$ time.

The whole construction of $A_{\text{root},i}$ takes $\mathcal{O}(z(T)\log^4 n) = \mathcal{O}(\delta(T)\log^5 n)$ time. Over all $i \in [4 .. \lceil \log n \rceil)$, this sums up to $\mathcal{O}(\delta(T)\log^6 n)$ time.

- 3. Next, we construct the structures for range queries. Let $i \in [4..\lceil \log n \rceil)$. Let g and T_{aux} be defined as above. Recall that above we observed that for every $j_1, j_2 \in \mathsf{R}(\tau_i, T)$, letting $s_1'' = \max \mathsf{Occ}_{\tau_i}(j_1, T_{\mathrm{aux}})$ and $s_2'' = \max \operatorname{Occ}_{\tau_i}(j_2, T_{\text{aux}}), g(s_1'' - n) = g(s_2'' - n)$ holds if and only if $\operatorname{root}_{f_i}(\tau_i, T, j_1) = \operatorname{root}_{f_i}(\tau_i, T, j_2)$. Since g(x) is of the form $1 + 2\tau_i \cdot x$, and for every $j \in \mathsf{R}(\tau_i, T)$, letting $s'' = \max \operatorname{Occ}_{\tau_i}(j, T_{\text{aux}})$ it holds $g(s''-n) \in [1..2\tau_i \cdot n_{\text{runs},i}]$, we thus obtain that the function $g': \mathsf{R}(\tau_i,T) \to [1..n_{\text{runs},i}]$ defined by $g'(j) = \frac{g(s''-n)-1}{2\tau_i} + 1$ (where s'' is defined as above) also uniquely identifies $\mathsf{root}_{f_i}(\tau_i,T,j)$, but returns a smaller integer. We thus proceed as follows:
 - (a) First, we repeat the computation from the previous step, except this time after we compute $s'' = \max \operatorname{Occ}_{\tau_i}(p_j, T_{\text{aux}})$ (where $j \in [1 ... n_{\text{runs},i}]$), we also store $g'(p_j) = \frac{g(s''-n)-1}{2\tau_i} + 1$ in an auxiliary array at index j. This takes $\mathcal{O}(\delta(T)\log^5 n)$ time.
 - (b) Next, we compute the collections $\operatorname{Seed}_{\tau_i,H}^-(\tau_i,T)$ (see Definition 8.60). To this end, we iterate over $A_{\text{root},i}[1..n_{\text{runs},i}]$. Let $(p,s) \stackrel{f_i,T}{=} A_{\text{root},i}[j] = (\text{head}_{f_i}(\tau_i,T,p_j),|\text{root}_{f_i}(\tau_i,T,p_j)|)$. Us- $\exp_{f_i}(\tau_i, T, p_j) \cdot p$. Next, using again CompSACore(T) in $\mathcal{O}(\log n)$ time we compute $b(\tau_i, T, p_j) =$ $p_i - \operatorname{lcs}(T[1 ... p_i), T[1 ... p_i + p))$. Lastly, we determine type (τ_i, T, p_i) by comparing $T[e(\tau_i, T, p_i)]$ with $T[e(\tau_i, T, p_j) - p]$ in $\mathcal{O}(\log n)$ time. Using the above values we can now calculate the values $\alpha(p_j) = e_{f_i}^{\text{full}}(\tau_i, T, p_j) \text{ and } \beta(p_j) = \min(7\tau_i, e_{f_i}^{\text{full}}(\tau_i, T, p_j) - b(\tau_i, T, p_j)). \text{ If } \text{type}(\tau_i, T, p_j) = -1,$ we store the resulting pair $(\alpha(p_j), \beta(p_j))$ in a linked list associated with position $g'(p_j)$ (which is retrieved from the auxiliary array computed above). Observe that since $g'(p_i)$ uniquely identifies $\operatorname{root}_{f_i}(\tau_i, T, p_j)$, it follows by Lemma 8.61 that the resulting linked lists contain precisely the

- collection $\{\text{Seed}_{f_i,H}^-(\tau_i,T)\}_{H\in\Sigma^+}$ (see Definition 8.60). Over all $j\in[1..n_{\text{runs},i}]$, the above process takes $\mathcal{O}(n_{\text{runs},i} \cdot \log n) = \mathcal{O}(\delta(T) \log n)$ time.
- (c) For each list we apply Proposition 6.6. Note that all queries needed in Proposition 6.6 are either supported using CompSACore(T) or using structures constructed above (in particular, we can lexicographically compare substrings using LCE and random access queries). The pointer to each structure is stored in a temporary array along with the pointer to the linked list containing each collection Seed $_{f_i,H}^-(\tau_i,T)$. Since the total length of lists is $n_{\text{runs},i}$, in total we thus spend $\mathcal{O}(n_{\text{runs}} \cdot \log^3 n) = \mathcal{O}(\delta(T) \log^3 n)$ time.
- (d) We perform a pass over the sequence $(p_j, t_j)_{j \in [1..n_{\text{runs},i}]}$. For each j, we retrieve the integer $g'(p_j)$ identifying $H = \text{root}_{f_i}(\tau_i, T, p_j)$, then obtain the pointer to the structure from Proposition 6.6 for $\operatorname{Seed}_{f_i,H}^-(\tau_i,T)$, and store it in $A_{\operatorname{ptr},i}[j]$. Note that this time we deliberately ignore $\operatorname{type}(\tau_i,T,p_j)$ (see Remark 8.80). This takes $\mathcal{O}(n_{\text{runs},i}) = \mathcal{O}(\delta(T))$ time.

In total the above computation takes $\mathcal{O}(\delta(T)\log^5 n)$ time. Over all $i \in [4..\lceil \log n \rceil)$, this sums up to $\mathcal{O}(\delta(T)\log^6 n)$ time.

4. We now compute the structures for weighted modular queries. Let $i \in [4.. \lceil \log n \rceil)$. Above we already constructed the collection of sets $\{\operatorname{Seed}_{f_i,H}^-(\tau_i,T)\}_{H\in\Sigma^+}$, each represented using a list. For each set Seed_{f_i,H}(τ_i,T), we apply Proposition 7.3 with $q=7\tau_i$ and h=|H|. Note again that all queries required by Proposition 7.3 are either supported using CompSACore(T) or using structures constructed above. Since the total length of lists is $n_{\text{runs},i}$, in total we thus spend $\mathcal{O}(n_{\text{runs},i} \cdot \log^3 n) = \mathcal{O}(\delta(T) \log^3 n)$ time. The pointer to each structure is stored in a temporary array along with the pointer to the linked list containing Seed $_{f_i,H}^-(\tau_i,T)$. After all structures are constructed, we again perform a pass over the sequence $(p_j, t_j)_{j \in [1...n_{\text{runs},i}]}$. For each j, we retrieve the integer $g'(p_j)$ identifying $H = \text{root}_{f_i}(\tau_i, T, p_j)$, then obtain the pointer to the structure from Proposition 7.3 for Seed $_{t_i,H}^-(\tau_i,T)$, and store it in $A'_{\text{ptr},i}[j]$. We again deliberately ignore type (τ_i, T, p_i) (see Remark 8.80). In total the above computation takes $\mathcal{O}(\delta(T)\log^3 n)$ time. Over all $i \in [4..[\log n])$, this sums up to $\mathcal{O}(\delta(T)\log^4 n)$ time.

In total, the construction of the first part of CompSAPeriodic(T) takes $\mathcal{O}(\delta(T)\log^7 n)$ time. We then construct the second part analogously. In total, the construction takes $\mathcal{O}(\delta(T)\log^7 n)$ time.

8.5 The Final Data Structure

The Data Structure 8.5.1

Definitions Let n_{short} be the number of distinct length-16 substrings of T^{∞} , and let Let $A_{\text{str}}[1...n_{\text{short}}]$ be an array containing all length-16 substring of T^{∞} sorted in lexicographic order. Let $A_{\rm range}[1\dots n_{\rm short}]$ be an array such that for every $i \in [1 \dots n_{\text{short}}]$, it holds $A_{\text{range}}[i] = (\text{RangeBeg}_{16}(A_{\text{str}}[i], T), \text{RangeEnd}_{16}(A_{\text{str}}[i], T))$. Finally, let $A_{\text{minocc}}[1 ... n_{\text{short}}]$ be such that for every $i \in [1 ... n_{\text{short}}]$, it holds $A_{\text{minocc}}[i] = \min \text{Occ}(A_{\text{str}}[i], T)$.

Components The data structure, denoted CompSA(T), consists of five parts:

- 1. The array $A_{\rm str}[1..n_{\rm short}]$ stored in plan form. Note that, by definition of $\delta(T)$, it holds $n_{\rm short} < 1$ $d_{16}(T) + 16 \le 16\delta(T) + 16 \le 32\delta(T)$. Thus, the array needs $\mathcal{O}(\delta(T))$ space.
- 2. The array $A_{\text{range}}[1..n_{\text{short}}]$ stored in plain form. It needs $\mathcal{O}(\delta(T))$ space.
- 3. The array $A_{\text{minocc}}[1..n_{\text{short}}]$ stored in plain form. It needs $\mathcal{O}(\delta(T))$ space.
- 4. The structure CompSANonperiodic(T) (Section 8.3.2). It needs $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ space. 5. The structure CompSAPeriodic(T) (Section 8.4.4). It needs $\mathcal{O}(\delta(T)\log\frac{n\log\sigma}{\delta(T)\log n})$ space.

In total, CompSA(T) needs $\mathcal{O}(\delta(T)\log \frac{n\log \sigma}{\delta(T)\log n})$ space.

Implementation of ISA Queries

Proposition 8.146. Let $j \in [1..n]$. Given CompSA(T) and the position j, we can compute the pair $(RangeBeg_{16}(j,T), RangeEnd_{16}(j,T))$ and a position $j' \in Occ_{16}(j,T)$ satisfying $j' = min Occ_{32}(j',T)$ in $\mathcal{O}(\log n)$ time.

Proof. In $\mathcal{O}(\log n)$ time we compute $X = T^{\infty}[j ... j + 16)$ using random access queries. Using binary search, in $\mathcal{O}(\log n)$ time we then compute $i \in [1...n_{\text{short}}]$ such that $A_{\text{str}}[i] = X$. Then,

```
\begin{split} A_{\text{range}}[i] &= (\text{RangeBeg}_{16}(A_{\text{str}}[i], T), \text{RangeEnd}_{16}(A_{\text{str}}[i], T)) \\ &= (\text{RangeBeg}_{16}(X, T), \text{RangeEnd}_{16}(X, T)) \\ &= (\text{RangeBeg}_{16}(j, T), \text{RangeEnd}_{16}(j, T)) \end{split}
```

and $A_{\text{minocc}}[i] = \min \text{Occ}_{16}(A_{\text{str}}[i], T) = \min \text{Occ}_{16}(X, T) = \min \text{Occ}_{16}(j, T)$. Letting $j' = A_{\text{minocc}}[i]$, we thus have $j' \in \text{Occ}_{16}(j, T)$. This in turn implies that $\text{Occ}_{16}(j, T) = \text{Occ}_{16}(j', T)$, and hence $j' = \min \text{Occ}_{16}(j, T) = \min \text{Occ}_{16}(j', T)$. It remains to observe that $\text{Occ}_{32}(j', T) \subseteq \text{Occ}_{16}(j', T)$. Thus, $j' = \min \text{Occ}_{32}(j', T)$.

Proposition 8.147. Let $k \in [4..\lceil \log n \rceil)$, $\ell = 2^k$, and $j \in [1..n]$. Given $\operatorname{CompSA}(T)$, the value k, the position j, the pair $(\operatorname{RangeBeg}_{\ell}(j,T), \operatorname{RangeEnd}_{\ell}(j,T))$, and a position $j' \in \operatorname{Occ}_{\ell}(j,T)$ satisfying $j' = \min \operatorname{Occ}_{2\ell}(j',T)$ as input, we can in $\mathcal{O}(\log^{3+\epsilon} n)$ time compute the pair $(\operatorname{RangeBeg}_{2\ell}(j,T), \operatorname{RangeEnd}_{2\ell}(j,T))$ and a position $j'' \in \operatorname{Occ}_{2\ell}(j,T)$ satisfying $j'' = \min \operatorname{Occ}_{4\ell}(j'',T)$.

Proof. Denote $\tau = \lfloor \frac{\ell}{3} \rfloor$. First, using CompSACore(T) and Proposition 8.17 in $\mathcal{O}(\log n)$ time we check if $j \in \mathbb{R}(\tau, T)$. We then consider two cases:

- If $j \notin R(\tau, T)$, then we obtain the output in $\mathcal{O}(\log^{2+\epsilon} n)$ time by applying Proposition 8.41.
- Otherwise $(j \in R(\tau, T))$, we obtain the output in $\mathcal{O}(\log^{3+\epsilon} n)$ time by applying Proposition 8.142.

In total, we spend $\mathcal{O}(\log^{3+\epsilon} n)$ time.

Proposition 8.148. Let $j \in [1..n]$. Given CompSA(T) and the position j as input, we can in $\mathcal{O}(\log^{4+\epsilon} n)$ time compute ISA[j].

Proof. First, using Proposition 8.146, in $\mathcal{O}(\log n)$ time we compute (RangeBeg₁₆(j,T), RangeEnd₁₆(j,T)) and a position $j' \in \operatorname{Occ}_{16}(j,T)$ satisfying $j' = \min \operatorname{Occ}_{32}(j',T)$. Then, for $k = 4, \ldots, \lceil \log n \rceil - 1$, we use Proposition 8.147 to compute in $\mathcal{O}(\log^{3+\epsilon} n)$ time (RangeBeg_{2k+1}(j,T), RangeEnd_{2k+1}(j,T)) and a position $j'' \in \operatorname{Occ}_{2k+1}(j,T)$ satisfying $j'' = \min \operatorname{Occ}_{2k+2}(j'',T)$, using position j along with the pair (RangeBeg_{2k}(j,T), RangeEnd_{2k}(j,T)) and $j' \in \operatorname{Occ}_{2k}(j,T)$ satisfying $j' = \min \operatorname{Occ}_{2k+1}(j',T)$ as input. After executing all steps, we obtain (RangeBeg₂(j,T), RangeEnd₂(j,T)) and a position $j'' \in \operatorname{Occ}_{\ell}(j,T)$, where $\ell = 2^{\lceil \log n \rceil} \geq n$. Since for every $\ell' \geq n$, it holds $\operatorname{Occ}_{\ell'}(j,T) = \{j\}$, we thus return $\operatorname{ISA}[j] = \operatorname{RangeEnd}_{\ell}(j,T)$ as the answer. In total, the query takes $\mathcal{O}(\log^{4+\epsilon} n)$ time.

8.5.3 Implementation of SA Queries

Proposition 8.149. Let $i \in [1..n]$. Given CompSA(T) and the position i, we can compute the pair $(RangeBeg_{16}(SA[i], T), RangeEnd_{16}(SA[i], T))$ and $j' \in Occ_{16}(SA[i], T)$ satisfying $j' = min Occ_{32}(j', T)$ in O(log n) time.

Proof. Using binary search, in $\mathcal{O}(\log n)$ time we compute $i' \in [1 ... n_{\text{short}}]$ such that, letting $A_{\text{range}}[t] = (b_t, e_t)$ (where $t \in [1 ... n_{\text{short}}]$), it holds $i \in (b_{i'} ... e_{i'}]$. Denote $X = A_{\text{str}}[i']$. We then have $T^{\infty}[\text{SA}[i] ... \text{SA}[i] + 16) = X$ and hence

```
\begin{split} A_{\text{range}}[i'] &= (\text{RangeBeg}_{16}(A_{\text{str}}[i'], T), \text{RangeEnd}_{16}(A_{\text{str}}[i'], T)) \\ &= (\text{RangeBeg}_{16}(X, T), \text{RangeEnd}_{16}(X, T)) \\ &= (\text{RangeBeg}_{16}(\text{SA}[i], T), \text{RangeEnd}_{16}(\text{SA}[i], T)). \end{split}
```

It also holds $A_{\text{minocc}}[i'] = \min \operatorname{Occ}_{16}(A_{\text{str}}[i'], T) = \min \operatorname{Occ}_{16}(X, T) = \min \operatorname{Occ}_{16}(\operatorname{SA}[i], T)$. Letting $j' = A_{\text{minocc}}[i']$, we thus have $j' \in \operatorname{Occ}_{16}(\operatorname{SA}[i], T)$. This in turn implies $\operatorname{Occ}_{16}(\operatorname{SA}[i], T) = \operatorname{Occ}_{16}(j', T)$, and hence $j' = \min \operatorname{Occ}_{16}(\operatorname{SA}[i], T) = \min \operatorname{Occ}_{16}(j', T)$. It remains to observe that $\operatorname{Occ}_{32}(j', T) \subseteq \operatorname{Occ}_{16}(j', T)$. Thus, $j' = \min \operatorname{Occ}_{32}(j', T)$.

Proposition 8.150. Let $k \in [4..\lceil \log n \rceil)$, $\ell = 2^k$, and $i \in [1..n]$. Given $\operatorname{CompSA}(T)$, the value k, the position i, the pair $(\operatorname{RangeBeg}_{\ell}(\operatorname{SA}[i],T),\operatorname{RangeEnd}_{\ell}(\operatorname{SA}[i],T))$, and $j' \in \operatorname{Occ}_{\ell}(\operatorname{SA}[i],T)$ satisfying $j' = \min \operatorname{Occ}_{2\ell}(j',T)$ as input, we can in $\mathcal{O}(\log^{3+\epsilon}n)$ time compute $(\operatorname{RangeBeg}_{2\ell}(\operatorname{SA}[i],T),\operatorname{RangeEnd}_{2\ell}(\operatorname{SA}[i],T))$ and a position $j' \in \operatorname{Occ}_{2\ell}(\operatorname{SA}[i],T)$ satisfying $j' = \min \operatorname{Occ}_{4\ell}(j',T)$.

Proof. Denote $\tau = \lfloor \frac{\ell}{3} \rfloor$. First, using CompSACore(T) and Proposition 8.17 in $\mathcal{O}(\log n)$ time we check if $j' \in \mathsf{R}(\tau, T)$. Observe that by $3\tau - 1 \leq \ell$ and $j' \in \mathsf{Occ}_{\ell}(\mathsf{SA}[i], T)$, $\mathsf{SA}[i] \in \mathsf{R}(\tau, T)$ holds if and only if $j' \in \mathsf{R}(\tau, T)$. Thus, this test reveals also whether $\mathsf{SA}[i] \in \mathsf{R}(\tau, T)$. We then consider two cases:

- If $SA[i] \notin R(\tau, T)$, then we compute the output using Proposition 8.42 in $\mathcal{O}(\log^{3+\epsilon} n)$ time.
- Otherwise (SA[i] $\in R(\tau, T)$), we compute the output using Proposition 8.144 in $\mathcal{O}(\log^{3+\epsilon} n)$ time.

In total, we spend $\mathcal{O}(\log^{3+\epsilon} n)$ time.

Proposition 8.151. Let $i \in [1..n]$. Given CompSA(T) and the position i as input, we can in $\mathcal{O}(\log^{4+\epsilon} n)$ time compute SA[i].

Proof. First, using Proposition 8.149, in $\mathcal{O}(\log n)$ time we compute the initial pair (RangeBeg₁₆(SA[i], T), RangeEnd₁₆(SA[i], T)) and a position $j' \in \operatorname{Occ}_{16}(\operatorname{SA}[i], T)$ satisfying $j' = \min \operatorname{Occ}_{32}(j', T)$. Then, for $k = 4, \ldots, \lceil \log n \rceil - 1$, we use Proposition 8.150 to compute in $\mathcal{O}(\log^{3+\epsilon} n)$ time the pair (RangeBeg_{2*+1}(SA[i], T), RangeEnd_{2*+1}(SA[i], T)) and a position $j'' \in \operatorname{Occ}_{2^{k+1}}(\operatorname{SA}[i], T)$ satisfying $j'' = \min \operatorname{Occ}_{2^{k+2}}(j'', T)$, using index i along with (RangeBeg_{2*}(SA[i], T), RangeEnd_{2*}(SA[i], T)) and $j' \in \operatorname{Occ}_{2^k}(\operatorname{SA}[i], T)$ satisfying $j' = \min \operatorname{Occ}_{2^{k+1}}(j', T)$ as input. After executing all steps, we obtain (RangeBeg₂(SA[i], T), RangeEnd_{\ell}(SA[i], T)) and a position $j'' \in \operatorname{Occ}_{\ell}(\operatorname{SA}[i], T)$, where $\ell = 2^{\lceil \log n \rceil} \geq n$. Since for every $\ell' \geq n$, it holds $\operatorname{Occ}_{\ell'}(\operatorname{SA}[i], T) = \{\operatorname{SA}[i]\}$, we thus return $\operatorname{SA}[i] = j''$ as the answer. In total, the query takes $\mathcal{O}(\log^{4+\epsilon} n)$ time.

8.5.4 Construction Algorithm

Proposition 8.152. Given the LZ77 parsing of T, we can construct CompSA(T) in $\mathcal{O}(\delta(T)\log^7 n)$ time.

Proof. First, using [KK20b, Theorem 6.11] (resp. [KK20b, Theorem 6.21]), in $\mathcal{O}(z(T)\log^4 n)$ time we construct a structure that, given any substring S of T (specified with its starting position and the length) in $\mathcal{O}(\log^3 n)$ time returns min $\operatorname{Occ}(S,T)$ (resp. $|\operatorname{Occ}(S,T)|$).

We then construct the components of CompSA(T) as follows:

- 1. Observe that every length-16 substring of T^{∞} overlaps a phrase boundary in the LZ77 parsing of T. In $\mathcal{O}(z(T)\log n)$ time we thus first construct an array containing all length-16 substrings of T^{∞} overlapping phrase boundaries. In $\mathcal{O}(z(T)\log z(T))$ time we then sort the array, and remove duplicates. The resulting array is precisely $A_{\rm str}$.
- 2. To compute A_{range} , we then proceed as follows. For $i = 1, \ldots, n_{\text{short}}$, letting $X_i = A_{\text{str}}[i]$, we compute $q_i = |\operatorname{Occ}(X_i, T)|$ in $\mathcal{O}(\log^3 n)$ time. If i = 1, we set $A_{\text{range}}[i] = (0, q_1)$. Otherwise, letting $(b, e) = A_{\text{range}}[i-1]$, we set $A_{\text{range}}[i] = (e, e+q_i)$. In total, we spend $\mathcal{O}(z(T)\log^3 n)$ time.
- 3. To compute A_{minocc} , we proceed as follows. For $i = 1, \ldots, n_{\text{short}}$, letting $X_i = A_{\text{str}}[i]$, we compute $m_i = \min \operatorname{Occ}(X_i, T)$ in $\mathcal{O}(\log^3 n)$ time, and then set $A_{\text{minocc}}[i] = m_i$. In total, we spend $\mathcal{O}(z(T)\log^3 n)$ time.
- 4. In $\mathcal{O}(\delta(T)\log^7 n)$ time we construct CompSANonperiodic(T) using Proposition 8.43.
- 5. In $\mathcal{O}(\delta(T)\log^7 n)$ time we construct CompSAPeriodic(T) using Proposition 8.145.

In total, the construction takes $\mathcal{O}(z(T)\log^3 n + \delta(T)\log^7 n) = \mathcal{O}(\delta(T)\log^7 n)$ time.

8.6 Summary

By combining Propositions 8.148, 8.151, and 8.152 and the size upper bound from Section 8.5.1, we obtain the following final result.

Theorem 1.1 (δ -SA). Given the LZ77 parsing of $T \in [0...\sigma)^n$ and any constant $\epsilon \in (0,1)$, we can in $\mathcal{O}(\delta \log^7 n)$ time construct a data structure of size $\mathcal{O}(\delta \log \frac{n \log \sigma}{\delta \log n})$ (where δ is the substring complexity of T) that, given any position $i \in [1..n]$, returns the values SA[i] and $SA^{-1}[i]$ in $\mathcal{O}(\log^{4+\epsilon} n)$ time. The construction algorithm is deterministic, and the running times are worst-case.

References

- [ABBK17] Amir Abboud, Arturs Backurs, Karl Bringmann, and Marvin Künnemann. Fine-grained complexity of analyzing compressed data: Quantifying improvements over decompress-and-solve. In Chris Umans, editor, 58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, pages 192–203. IEEE Computer Society, 2017. doi:10.1109/FOCS.2017.26.
- [ABBK20] Amir Abboud, Arturs Backurs, Karl Bringmann, and Marvin Künnemann. Impossibility results for grammar-compressed linear algebra. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, 34th Conference on Neural Information Processing System, NeurIPS 2020, volume 33, pages 8810–8823. Curran Associates, Inc., 2020. URL: https://proceedings.neurips.cc/paper/2020/file/645e6bfdd05d1a69c5e47b20f0a91d46-Paper.pdf.
- [ABM08] Donald Adjeroh, Tim Bell, and Amar Mukherjee. *The Burrows-Wheeler Transform: Data Compression, Suffix Arrays, and Pattern Matching.* Springer, Boston, MA, USA, 2008. doi: 10.1007/978-0-387-78909-5.
- [AJ22] Shyan Akmal and Ce Jin. Near-optimal quantum algorithms for string problems. In Joseph (Seffi) Naor and Niv Buchbinder, editors, 33rd Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, pages 2791–2832. SIAM, 2022. doi:10.1137/1.9781611977073.109.
- [BCG⁺21] Djamal Belazzougui, Manuel Cáceres, Travis Gagie, Paweł Gawrychowski, Juha Kärkkäinen, Gonzalo Navarro, Alberto Ordóñez Pereira, Simon J. Puglisi, and Yasuo Tabei. Block trees. Journal of Computer and System Sciences, 117:1–22, 2021. doi:10.1016/j.jcss.2020.11.002.
- [BDY16] Bonnie Berger, Noah M. Daniels, and Y. William Yu. Computational biology in the 21st century: Scaling with compressive algorithms. *Communications of the ACM*, 59(8):72–80, 2016. doi:10.1145/2957324.
- [Bel14] Djamal Belazzougui. Linear time construction of compressed text indices in compact space. In David B. Shmoys, editor, 46th Annual ACM Symposium on Theory of Computing, STOC 2014, pages 148–193. ACM, 2014. doi:10.1145/2591796.2591885.
- [BKW19] Karl Bringmann, Marvin Künnemann, and Philip Wellnitz. Few matches or almost periodicity: Faster pattern matching with mismatches in compressed texts. In Timothy M. Chan, editor, 30th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, pages 1126–1145. SIAM, 2019. doi:10.1137/1.9781611975482.69.
- [BLR⁺15] Philip Bille, Gad M. Landau, Rajeev Raman, Kunihiko Sadakane, Srinivasa Rao Satti, and Oren Weimann. Random access to grammar-compressed strings and trees. *SIAM Journal on Computing*, 44(3):513–539, 2015. doi:10.1137/130936889.
- [BN14] Djamal Belazzougui and Gonzalo Navarro. Alphabet-independent compressed text indexing. ACM Transactions on Algorithms, 10(4):23:1–23:19, 2014. doi:10.1145/2635816.
- [BW94] Michael Burrows and David J. Wheeler. A block-sorting lossless data compression algorithm. Technical Report 124, Digital Equipment Corporation, Palo Alto, California, 1994. URL: https://www.hpl.hp.com/techreports/Compaq-DEC/SRC-RR-124.pdf.

- [CEK⁺21] Anders Roy Christiansen, Mikko Berggren Ettienne, Tomasz Kociumaka, Gonzalo Navarro, and Nicola Prezza. Optimal-time dictionary-compressed indexes. *ACM Transactions on Algorithms*, 17(1):8:1–8:39, 2021. doi:10.1145/3426473.
- [Cha88] Bernard Chazelle. A functional approach to data structures and its use in multidimensional searching. SIAM Journal on Computing, 17(3):427–462, 1988. doi:10.1137/0217026.
- [CHL07] Maxime Crochemore, Christophe Hancart, and Thierry Lecroq. Algorithms on strings. Cambridge University Press, Cambridge, UK, 2007. doi:10.1017/cbo9780511546853.
- [CKPR21] Panagiotis Charalampopoulos, Tomasz Kociumaka, Solon P. Pissis, and Jakub Radoszewski. Faster algorithms for longest common substring. In Petra Mutzel, Rasmus Pagh, and Grzegorz Herman, editors, 29th Annual European Symposium on Algorithms, ESA 2021, volume 204 of LIPIcs, pages 30:1–30:17. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021. doi: 10.4230/LIPIcs.ESA.2021.30.
- [CKW20] Panagiotis Charalampopoulos, Tomasz Kociumaka, and Philip Wellnitz. Faster approximate pattern matching: A unified approach. In Sandy Irani, editor, 61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, pages 978–989. IEEE Computer Society, 2020. doi:10.1109/F0CS46700.2020.00095.
- [CLL⁺05] Moses Charikar, Eric Lehman, Ding Liu, Rina Panigrahy, Manoj Prabhakaran, Amit Sahai, and Abhi Shelat. The smallest grammar problem. *IEEE Transactions on Information Theory*, 51(7):2554–2576, 2005. doi:10.1109/TIT.2005.850116.
- [CN11] Francisco Claude and Gonzalo Navarro. Self-indexed grammar-based compression. Fundamenta Informaticae, 111(3):313–337, 2011. doi:10.3233/FI-2011-565.
- [CNP21] Francisco Claude, Gonzalo Navarro, and Alejandro Pacheco. Grammar-compressed indexes with logarithmic search time. *Journal of Computer and System Sciences*, 118:53–74, 2021. doi:10.1016/j.jcss.2020.12.001.
- [DFH⁺20] Patrick Dinklage, Johannes Fischer, Alexander Herlez, Tomasz Kociumaka, and Florian Kurpicz. Practical performance of space efficient data structures for longest common extensions. In Fabrizio Grandoni, Grzegorz Herman, and Peter Sanders, editors, 28th Annual European Symposium on Algorithms, ESA 2020, volume 173 of LIPIcs, pages 39:1–39:20. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ESA.2020.39.
- [DNP21] Diego Díaz-Domínguez, Gonzalo Navarro, and Alejandro Pacheco. An LMS-based grammar self-index with local consistency properties. In Thierry Lecroq and Hélène Touzet, editors, 28th International Symposium on String Processing and Information Retrieval SPIRE 2021, volume 12944 of LNCS, pages 100–113. Springer, 2021. doi:10.1007/978-3-030-86692-1_9.
- [FGK⁺22] Alexandre P. Francisco, Travis Gagie, Dominik Köppl, Susana Ladra, and Gonzalo Navarro. Graph compression for adjacency-matrix multiplication. *SN Computer Science*, 3(3):193, 2022. doi:10.1007/s42979-022-01084-2.
- [FM05] Paolo Ferragina and Giovanni Manzini. Indexing compressed text. *Journal of the ACM*, 52(4):552–581, 2005. doi:10.1145/1082036.1082039.
- [FM10] Paolo Ferragina and Giovanni Manzini. On compressing the textual web. In Brian D. Davison, Torsten Suel, Nick Craswell, and Bing Liu, editors, 3rd International Conference on Web Search and Web Data Mining, WSDM 2010, pages 391–400. ACM, 2010. doi: 10.1145/1718487.1718536.

- [FMG⁺22] Paolo Ferragina, Giovanni Manzini, Travis Gagie, Dominik Köppl, Gonzalo Navarro, Manuel Striani, and Francesco Tosoni. Improving matrix-vector multiplication via lossless grammar-compressed matrices. *Proceedings of the VLDB Endowment*, 15(10):2175–2187, 2022. URL: https://www.vldb.org/pvldb/vol15/p2175-tosoni.pdf.
- [FW65] Nathan J. Fine and Herbert S. Wilf. Uniqueness theorems for periodic functions. *Proceedings of the American Mathematical Society*, 16(1):109–114, 1965. doi:10.2307/2034009.
- [Gaw11] Paweł Gawrychowski. Pattern matching in Lempel-Ziv compressed strings: Fast, simple, and deterministic. In Camil Demetrescu and Magnús M. Halldórsson, editors, 19th Annual European Symposium on Algorithms, ESA 2011, volume 6942 of LNCS, pages 421–432. Springer, 2011. doi:10.1007/978-3-642-23719-5_36.
- [Gaw13] Paweł Gawrychowski. Optimal pattern matching in LZW compressed strings. ACM Transactions on Algorithms, 9(3):25:1–25:17, 2013. doi:10.1145/2483699.2483705.
- [GBMP14] Simon Gog, Timo Beller, Alistair Moffat, and Matthias Petri. From theory to practice: Plug and play with succinct data structures. In Joachim Gudmundsson and Jyrki Katajainen, editors, 13th International Symposium on Experimental Algorithms, SEA 2014, volume 8504 of LNCS, pages 326–337. Springer, 2014. doi:10.1007/978-3-319-07959-2_28.
- [GG22] Moses Ganardi and Paweł Gawrychowski. Pattern matching on grammar-compressed strings in linear time. In Joseph (Seffi) Naor and Niv Buchbinder, editors, 33rd Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, pages 2833–2846. SIAM, 2022. doi:10.1137/1.9781611977073.110.
- [GGK+12] Travis Gagie, Paweł Gawrychowski, Juha Kärkkäinen, Yakov Nekrich, and Simon J. Puglisi. A faster grammar-based self-index. In Adrian-Horia Dediu and Carlos Martín-Vide, editors, 6th International Conference on Language and Automata Theory and Applications, LATA 2012, volume 7183 of LNCS, pages 240–251. Springer, 2012. doi:10.1007/978-3-642-28332-1_21.
- [GGK⁺14] Travis Gagie, Paweł Gawrychowski, Juha Kärkkäinen, Yakov Nekrich, and Simon J. Puglisi. LZ77-based self-indexing with faster pattern matching. In Alberto Pardo and Alfredo Viola, editors, 11th Latin American Symposium on Theoretical Informatics, LATIN 2014, volume 8392 of LNCS, pages 731–742. Springer, 2014. doi:10.1007/978-3-642-54423-1_63.
- [GGV03] Roberto Grossi, Ankur Gupta, and Jeffrey Scott Vitter. High-order entropy-compressed text indexes. In 14th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2003, pages 841–850. ACM/SIAM, 2003. URL: http://dl.acm.org/citation.cfm?id=644108.644250.
- [GJL21] Moses Ganardi, Artur Jeż, and Markus Lohrey. Balancing straight-line programs. *Journal of the ACM*, 68(4):27:1–27:40, 2021. doi:10.1145/3457389.
- [GKK⁺18] Paweł Gawrychowski, Adam Karczmarz, Tomasz Kociumaka, Jakub Łącki, and Piotr Sankowski. Optimal dynamic strings. In Artur Czumaj, editor, 29th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, pages 1509–1528. SIAM, 2018. doi:10.1137/1.9781611975031.
- [GKLS22] Arun Ganesh, Tomasz Kociumaka, Andrea Lincoln, and Barna Saha. How compression and approximation affect efficiency in string distance measures. In Joseph (Seffi) Naor and Niv Buchbinder, editors, 33rd Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, pages 2867–2919. SIAM, 2022. doi:10.1137/1.9781611977073.112.
- [GNP18] Travis Gagie, Gonzalo Navarro, and Nicola Prezza. On the approximation ratio of Lempel-Ziv parsing. In Michael A. Bender, Martin Farach-Colton, and Miguel A. Mosteiro, editors, 13th Latin American Symposium on Theoretical Informatics, LATIN 2018, volume 10807 of LNCS, pages 490–503. Springer, 2018. doi:10.1007/978-3-319-77404-6_36.

- [GNP20] Travis Gagie, Gonzalo Navarro, and Nicola Prezza. Fully functional suffix trees and optimal text searching in BWT-runs bounded space. *Journal of the ACM*, 67(1):2:1–2:54, 2020. doi: 10.1145/3375890.
- [GT23] Daniel Gibney and Sharma V. Thankachan. Compressibility-aware quantum algorithms on strings, 2023. doi:10.48550/ARXIV.2302.07235.
- [Gus97] Dan Gusfield. Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology. Cambridge University Press, Cambridge, UK, 1997. doi:10.1017/cbo9780511574931.
- [GV05] Roberto Grossi and Jeffrey Scott Vitter. Compressed suffix arrays and suffix trees with applications to text indexing and string matching. SIAM Journal on Computing, 35(2):378–407, 2005. doi:10.1137/S0097539702402354.
- [Hag98] Torben Hagerup. Sorting and searching on the word RAM. In Michel Morvan, Christoph Meinel, and Daniel Krob, editors, 15th Annual Symposium on Theoretical Aspects of Computer Science, STACS 1998, volume 1373 of LNCS, pages 366–398. Springer, 1998. doi:10.1007/BFb0028575.
- [HLLW13] Danny Hermelin, Gad M. Landau, Shir Landau, and Oren Weimann. Unified compression-based acceleration of edit-distance computation. *Algorithmica*, 65(2):339–353, 2013. doi:10.1007/s00453-011-9590-6.
- [HPWO19] Mikel Hernaez, Dmitri Pavlichin, Tsachy Weissman, and Idoia Ochoa. Genomic data compression. Annual Review of Biomedical Data Science, 2(1):19–37, 2019. doi:10.1146/annurev-biodatasci-072018-021229.
- [HSS09] Wing-Kai Hon, Kunihiko Sadakane, and Wing-Kin Sung. Breaking a time-and-space barrier in constructing full-text indices. SIAM Journal on Computing, 38(6):2162–2178, 2009. doi: 10.1137/070685373.
- [I17] Tomohiro I. Longest common extensions with recompression. In Juha Kärkkäinen, Jakub Radoszewski, and Wojciech Rytter, editors, 28th Annual Symposium on Combinatorial Pattern Matching, CPM 2017, volume 78 of LIPIcs, pages 18:1–18:15. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2017. doi:10.4230/LIPIcs.CPM.2017.18.
- [Jeż15] Artur Jeż. Faster fully compressed pattern matching by recompression. *ACM Transactions on Algorithms*, 11(3):20:1–20:43, 2015. doi:10.1145/2631920.
- [JN23] Ce Jin and Jakob Nogler. Quantum speed-ups for string synchronizing sets, longest common substring, and k-mismatch matching. In Nikhil Bansal and Viswanath Nagarajan, editors, 34th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, pages 5090–5121. SIAM, 2023. doi:10.1137/1.9781611977554.ch186.
- [Kem19] Dominik Kempa. Optimal construction of compressed indexes for highly repetitive texts. In Timothy M. Chan, editor, 30th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, pages 1344–1357. SIAM, 2019. doi:10.1137/1.9781611975482.82.
- [KK99] Roman M. Kolpakov and Gregory Kucherov. Finding maximal repetitions in a word in linear time. In 40th IEEE Annual Symposium on Foundations of Computer Science, FOCS 1999, pages 596–604. IEEE Computer Society, 1999. doi:10.1109/SFFCS.1999.814634.
- [KK19] Dominik Kempa and Tomasz Kociumaka. String synchronizing sets: Sublinear-time BWT construction and optimal LCE data structure. In Moses Charikar and Edith Cohen, editors, 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, pages 756–767. ACM, 2019. doi:10.1145/3313276.3316368.

- [KK20a] Dominik Kempa and Tomasz Kociumaka. Resolution of the Burrows-Wheeler Transform conjecture. In Sandy Irani, editor, 61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, pages 1002–1013. IEEE Computer Society, 2020. doi:10.1109/F0CS46700.2020.00097.
- [KK20b] Dominik Kempa and Tomasz Kociumaka. Resolution of the Burrows-Wheeler Transform conjecture, 2020. Full version of [KK20a]. arXiv:1910.10631v3.
- [KK22] Dominik Kempa and Tomasz Kociumaka. Dynamic suffix array with polylogarithmic queries and updates. In Stefano Leonardi and Anupam Gupta, editors, 54th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2022, pages 1657–1670. ACM, 2022. doi:10.1145/3519935.3520061.
- [KK23] Dominik Kempa and Tomasz Kociumaka. Breaking the O(n)-barrier in the construction of compressed suffix arrays and suffix trees. In Nikhil Bansal and Viswanath Nagarajan, editors, 34th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, pages 5122–5202. SIAM, 2023. doi:10.1137/1.9781611977554.ch187.
- [KMS⁺03] Takuya Kida, Tetsuya Matsumoto, Yusuke Shibata, Masayuki Takeda, Ayumi Shinohara, and Setsuo Arikawa. Collage system: A unifying framework for compressed pattern matching. Theoretical Computer Science, 298(1):253–272, 2003. doi:10.1016/S0304-3975(02)00426-7.
- [KN10] Sebastian Kreft and Gonzalo Navarro. LZ77-like compression with fast random access. In 2010 Data Compression Conference, pages 239–248. IEEE Computer Society, 2010. doi:10.1109/DCC. 2010.29.
- [KNO22] Tomasz Kociumaka, Gonzalo Navarro, and Francisco Olivares. Near-optimal search time in δ-optimal space. In Armando Castañeda and Francisco Rodríguez-Henríquez, editors, 15th Latin American Symposium on Theoretical Informatics, LATIN 2022, volume 13568 of LNCS, pages 88–103. Springer, 2022. doi:10.1007/978-3-031-20624-5_6.
- [KNP23] Tomasz Kociumaka, Gonzalo Navarro, and Nicola Prezza. Towards a definitive compressibility measure for repetitive sequences. *IEEE Transactions on Information Theory*, 69(4):2074–2092, 2023. doi:10.1109/TIT.2022.3224382.
- [Koc18] Tomasz Kociumaka. Efficient Data Structures for Internal Queries in Texts. PhD thesis, University of Warsaw, 2018. URL: https://www.mimuw.edu.pl/~kociumaka/files/phd.pdf.
- [KP18] Dominik Kempa and Nicola Prezza. At the roots of dictionary compression: String attractors. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, pages 827–840. ACM, 2018. doi:10.1145/ 3188745.3188814.
- [KRRW23] Tomasz Kociumaka, Jakub Radoszewski, Wojciech Rytter, and Tomasz Waleń. Internal pattern matching queries in a text and applications, 2023. arXiv:1311.6235v5.
- [KS22] Dominik Kempa and Barna Saha. An upper bound and linear-space queries on the LZ-end parsing. In Joseph (Seffi) Naor and Niv Buchbinder, editors, 33rd Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, pages 2847–2866. SIAM, 2022. doi:10.1137/1.9781611977073. 111.
- [KVNP20] Dmitry Kosolobov, Daniel Valenzuela, Gonzalo Navarro, and Simon J. Puglisi. Lempel-Ziv-like parsing in small space. Algorithmica, 82(11):3195–3215, 2020. doi:10.1007/s00453-020-00722-6.
- [KY00] John C. Kieffer and En-Hui Yang. Grammar-based codes: A new class of universal lossless source codes. *IEEE Transactions on Information Theory*, 46(3):737–754, 2000. doi:10.1109/18.841160.

- [LZ76] Abraham Lempel and Jacob Ziv. On the complexity of finite sequences. *IEEE Transactions on Information Theory*, 22(1):75–81, 1976. doi:10.1109/TIT.1976.1055501.
- [Mai89] Michael G. Main. Detecting leftmost maximal periodicities. Discrete Applied Mathematics, 25(1-2):145–153, 1989. doi:10.1016/0166-218X(89)90051-6.
- [MNN17] J. Ian Munro, Gonzalo Navarro, and Yakov Nekrich. Space-efficient construction of compressed indexes in deterministic linear time. In Philip N. Klein, editor, 28th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, pages 408–424. SIAM, 2017. doi:10.1137/1.9781611974782.26.
- [Nat22] National Human Genome Research Institute (NIH). Genomic data science, 2022. Accessed March 30, 2023. URL: https://www.genome.gov/about-genomics/fact-sheets/Genomic-Data-Science.
- [Nav14] Gonzalo Navarro. Wavelet trees for all. Journal of Discrete Algorithms, 25:2-20, 2014. doi: 10.1016/j.jda.2013.07.004.
- [Nav16] Gonzalo Navarro. Compact data structures: A practical approach. Cambridge University Press, Cambridge, UK, 2016. doi:10.1017/cbo9781316588284.
- [Nav21a] Gonzalo Navarro. Indexing highly repetitive string collections, part I: Repetitiveness measures. ACM Computing Surveys, 54(2):29:1–29:31, 2021. doi:10.1145/3434399.
- [Nav21b] Gonzalo Navarro. Indexing highly repetitive string collections, part II: Compressed indexes. *ACM Computing Surveys*, 54(2):26:1–26:32, 2021. doi:10.1145/3432999.
- [NII+16] Takaaki Nishimoto, Tomohiro I, Shunsuke Inenaga, Hideo Bannai, and Masayuki Takeda. Fully dynamic data structure for LCE queries in compressed space. In Piotr Faliszewski, Anca Muscholl, and Rolf Niedermeier, editors, 41st International Symposium on Mathematical Foundations of Computer Science, MFCS 2016, volume 58 of LIPIcs, pages 72:1–72:15. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPIcs.MFCS.2016.72.
- [NM07] Gonzalo Navarro and Veli Mäkinen. Compressed full-text indexes. *ACM Computing Surveys*, 39(1):2:1–2:61, 2007. doi:10.1145/1216370.1216372.
- [PHR00] Molly Przeworski, Richard R. Hudson, and Anna Di Rienzo. Adjusting the focus on human variation. Trends in Genetics, 16(7):296–302, 2000. doi:10.1016/S0168-9525(00)02030-8.
- [PNB17] Alberto Ordóñez Pereira, Gonzalo Navarro, and Nieves R. Brisaboa. Grammar compressed sequences with rank/select support. *Journal of Discrete Algorithms*, 43:54–71, 2017. doi: 10.1016/j.jda.2016.10.001.
- [Pre19] Nicola Prezza. Optimal rank and select queries on dictionary-compressed text. In Nadia Pisanti and Solon P. Pissis, editors, 30th Annual Symposium on Combinatorial Pattern Matching, CPM 2019, volume 128 of LIPIcs, pages 4:1–4:12. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.CPM.2019.4.
- [Ryt03] Wojciech Rytter. Application of Lempel–Ziv factorization to the approximation of grammar-based compression. *Theoretical Computer Science*, 302(1–3):211–222, 2003. doi:10.1016/S0304-3975(02)00777-6.
- [Sad02] Kunihiko Sadakane. Succinct representations of LCP information and improvements in the compressed suffix arrays. In 13th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2002, pages 225–232. ACM/SIAM, 2002. URL: http://dl.acm.org/citation.cfm?id=545381.545410.

- [SS82] James A. Storer and Thomas G. Szymanski. Data compression via textual substitution. *Journal of the ACM*, 29(4):928–951, 1982. doi:10.1145/322344.322346.
- [Tis15] Alexander Tiskin. Fast distance multiplication of unit-Monge matrices. *Algorithmica*, 71(4):859–888, 2015. doi:10.1007/s00453-013-9830-z.
- [ZL77] Jacob Ziv and Abraham Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3):337–343, 1977. doi:10.1109/TIT.1977.1055714.