

# AUTOMATIC SEQUENCES FULFILL THE SARNAK CONJECTURE

CLEMENS MÜLLNER

ABSTRACT. We present in this paper a new method to deal with automatic sequences. This method allows us to prove a Möbius-randomness-principle for automatic sequences from which we deduce the Sarnak conjecture for this class of sequences. Furthermore, we can show a Prime Number Theorem for automatic sequences that are generated by strongly connected automata where the initial state is fixed by the transition corresponding to 0.

## 1. INTRODUCTION

In 2009, Peter Sarnak stated the following conjecture [33]:

**Conjecture 1.1.** *Let  $\mu$  be the Möbius function. For any bounded sequence  $\xi(n)$  observed by a deterministic flow <sup>1</sup>  $(X, S)$ , it holds that*

$$\sum_{n \leq N} \xi(n) \mu(n) = o(N).$$

This conjecture has aroused great interest and numerous papers were recently devoted to show the Sarnak conjecture for different classes of flows [2, 3, 4, 6, 8, 11, 12, 15, 19, 20, 22, 23, 25, 28, 29, 32, 34, 38].

Furthermore, we would like to mention two results connecting the Sarnak conjecture to the Chowla conjecture [17, 37].

Sarnak stated his conjecture in terms of a deterministic flow  $(X, S)$ . We are interested in special flows that origin from a sequence:

Fix a sequence  $\mathbf{a} = (a_n)_{n \in \mathbb{N}}$  that takes values in a finite set  $\mathcal{A}$ . We denote by  $S$  the shift operator on the sequences  $(\mathcal{A}^{\mathbb{N}})$  in  $\mathcal{A}$  and define  $X := \overline{\{S^n(\mathbf{a}) : n \in \mathbb{N}_0\}}$ . By taking the product topology of the discrete topologies on each copy of  $\mathcal{A}$ , we find that  $\mathcal{A}^{\mathbb{N}}$  is compact and complete. The product topology is induced by the following metric  $d(x, y) = \sum_{n=0}^{\infty} 2^{-n-1} d_n(x_n, y_n)$  on  $X$ , where  $d_n$  denotes the discrete metric on  $\mathcal{A}$ . The dynamical system (or flow)  $(X, S)$  is called the *symbolic dynamical system associated with  $\mathbf{a}$* .

---

This research is supported by the project F55-02 of the Austrian Science Fund FWF which is part of the Special Research Program “Quasi-Monte Carlo Methods: Theory and Applications”, by Project F5002-N15 (FWF), which is a part of the Special Research Program “Algorithmic and Enumerative Combinatorics” and by Project I1751 (FWF), called MUDERA (Multiplicativity, Determinism, and Randomness). Furthermore, the author wants to thank Michael Drmota for introducing this topic as well as for his help and guidance.

<sup>1</sup>A flow  $F$  is a pair  $(X, S)$  where  $X$  is a compact metric space and  $S : X \rightarrow X$  is a continuous map. For a dynamical system to be deterministic means that - roughly speaking - there are only few different orbits.

Therefore, we say that a sequence  $\mathbf{a}$  fulfills the Sarnak conjecture if the symbolic dynamical system associated with  $\mathbf{a}$  fulfills the Sarnak conjecture. For more information about symbolic dynamical systems and their complexity see for example [13, 16].

A Möbius-randomness principle is often closely related to a Prime Number Theorem (PNT), i.e., an asymptotic formula for the sum  $\sum_{n \leq x} \Lambda(n) u_n$ . Some classes or examples of automatic sequences, which are observed by a deterministic flow, have already been covered:

- the Thue-Morse sequence:
  - a Möbius-randomness principle follows from the work of Indlekofer and Kátai [24, 26]
  - Dartyge and Tenenbaum [5] give additionally an explicit error bound
  - a Möbius-randomness principle and a PNT by Mauduit and Rivat [30]
- the Rudin-Shapiro sequence:
  - Tao suggests a strategy to prove a Möbius-randomness principle in [36]
  - Mauduit and Rivat provide a different approach which allows to prove a Möbius-randomness principle and a PNT for more general functions with an explicit error bound [31]
- sequences generated by invertible automata
  - Sarnak conjecture [14]
  - a Möbius-randomness principle and a PNT [9]
- sequences generated by synchronizing automata
  - the Sarnak conjecture and a PNT [7]
- some special digital functions (e.g., the parity of occurrences of a word in the digital expansion)
  - a Möbius-randomness principle together with a PNT [21]

The purpose of this paper is to show the Sarnak conjecture for all automatic sequences, which covers all results concerning automatic sequences mentioned above.

**Theorem 1.2.** *Let  $\mu$  be the Möbius function,  $(a_n)_{n \in \mathbb{N}}$  be a complex valued automatic sequence and let  $(X, S)$  be the symbolic dynamical system associated with  $(a_n)_{n \in \mathbb{N}}$ . Then for all sequences  $\xi(n) := f(S^n(x))$ , with  $x \in X$  and  $f \in C(X, \mathbb{C})$ , we have*

$$\sum_{n \leq N} \xi(n) \mu(n) = o(N).$$

**Remark.** *This theorem does not provide an explicit error term. This is mainly due to the fact that we consider the symbolic dynamical system associated with an automatic sequence. When considering only  $\xi(n) = a_n$  we have basically two contributions for the error term:*

- *Estimates for the Möbius function along arithmetic progressions.*
- *Estimates provided by Theorem 4.4 which gives a power saving.*

*However, for the general case we use Lemma 3.1 to relate  $\xi(n)$  to other automatic sequences which provides no explicit error term.*

We present in this paper a new method which relates the output of a strongly connected deterministic finite automaton to the output of a newly introduced special class of transducers (so called *naturally induced transducers*). These transducers combine synchronizing and



One finds easily that the automatic sequence is equally distributed on  $\{b, c\}$ , as the generating automaton has only one final component which is primitive and symmetric. But, if we consider only odd integers we are detecting the first digit of  $n$  in base 3: the subsequence of odd integers of the automatic sequence is constant on all intervals of the form  $[3^m, 2 \cdot 3^m)$ ,  $[2 \cdot 3^m, 3^{m+1})$ . This shows that the densities of  $b$  and  $c$  do not exist for the subsequence of odd integers. This consideration together with the prime number theorem also shows that the densities of  $b$  and  $c$  do not exist along the primes.

In Section 2, we properly define naturally induced transducers. Furthermore, we describe a way to construct them and study some of their properties. In Section 3, we show how to use the concept of (naturally) induced transducers to reduce Theorem 1.2 and Theorem 1.3 to corresponding results on naturally induced transducers. To actually show these results, we use some modified results of [31]. This is done in Section 4. Thereafter, we show that the output of a naturally induced transducer fulfills a carry lemma and has uniformly small Fourier-terms. All the more technical proofs are postponed to the last Section 5.

**1.1. Definitions.** In this paper we let  $\mathbb{N}$  denote the set of positive integers, by  $\mathbb{U}$  the set of complex numbers of modulus 1 and we use the abbreviation  $e(x) = \exp(2\pi i x)$  for any real number  $x$ .

For two functions,  $f$  and  $g$ , that take only strictly positive real values such that  $f/g$  is bounded, we write  $f = O(g)$  or  $f \ll g$ .

We let  $\lfloor x \rfloor$  denote the floor function and  $\{x\}$  denote the fractional part of  $x$ . Furthermore, we let  $\chi_\alpha(x)$  denote the indicator function for  $\{x\}$  in  $[0, \alpha)$ .

Moreover we let  $\tau(n)$  denote the number of divisors of  $n$ ,  $\omega(n)$  denote the number of distinct prime factors of  $n$ ,  $\varphi(n)$  denote the number of integers smaller than  $n$  that are co-prime to  $n$  and  $\mathcal{P}(a, m)$  denote the set of prime numbers  $p \equiv a \pmod{m}$ .

Next we give all definitions related to automata which can also be found in [1].

**Definition 1.4.** A deterministic finite automaton, or DFA, is a quadruple  $A = (Q', \Sigma, \delta', q'_0)$ , where  $Q'$  is a finite set of states,  $\Sigma$  is the finite input alphabet,  $\delta' : Q' \times \Sigma \rightarrow Q'$  is the transition function and  $q'_0 \in Q'$  is the initial state.

We extend  $\delta$  to a function  $\delta : Q \times \Sigma^* \rightarrow Q$ , in a natural way, by

$$\delta(q, ab) = \delta(\delta(q, a), b)$$

for all  $a, b \in \Sigma^*$ . By definition  $\delta(q, \mathbf{w})$  consists of  $|\mathbf{w}|$  “steps” for every  $\mathbf{w} \in \Sigma^*$ .

We usually work with  $\mathbf{w} \in \Sigma^* = \{0, \dots, k-1\}^*$ , where we let  $\Sigma^*$  denote the set of all (finite) words over the alphabet  $\Sigma$ . We can view  $\mathbf{w}$  as the digital representation of a natural number  $n$  in base  $k$ . We define for  $\mathbf{w} = (w_0, \dots, w_r)$  the corresponding natural number  $[\mathbf{w}]_k := \sum_{i=0}^r k^{r-i} w_i$  (i.e., we assume that the first digit is the most significant). We denote for a natural number  $n$  the corresponding digit representation in base  $k$  (without leading zeros) by  $(n)_k$ . Furthermore, we let  $(n)_k^t$  denote the unique word  $\mathbf{w}$  of length  $t$  such that  $[\mathbf{w}]_k \equiv n \pmod{k^t}$ .

**Example.** We find  $(37)_2 = (1, 0, 0, 1, 0, 1)$ ,  $(37)_2^4 = (0, 1, 0, 1)$  and  $[(1, 0, 1, 1, 0)]_2 = 22$ .

As already mentioned, we always read words from left to right i.e., for digital representations of numbers we start with the most significant digit.

**Definition 1.5.** A DFAO  $A = (Q', \Sigma, \delta', q'_0, \Delta, \tau)$  is a DFA with an additional output function  $\tau : Q' \rightarrow \Delta$ .

**Definition 1.6.** We say that a sequence  $\mathbf{a} = (a_n)_{n \in \mathbb{N}}$  is a  $k$ -automatic sequence, if and only if there exists a DFAO  $A = (Q, \Sigma = \{0, \dots, k-1\}, \delta, q_0, \Delta, \tau)$  such that  $a_n = \tau(\delta(q_0, (n)_k))$ .

In this paper we restrict ourself to complex valued output functions; i.e.,  $\tau : Q' \rightarrow \mathbb{C}$  and we omit  $\Delta = \mathbb{C}$  from now on. Note that we can consider in Theorem 1.3 without loss of generality only complex valued automatic sequences.

We call an automaton (with output)  $A = (Q', \Sigma, \delta', q'_0)$  strongly connected if and only if for all  $q'_1, q'_2 \in Q$  there exists  $\mathbf{w} \in \Sigma^*$  with  $\delta'(q'_1, \mathbf{w}) = q'_2$ . We denote by a final component a strongly connected component that is closed under  $\delta'$ .

Furthermore, we let  $p(A)$  denote the period of an automaton  $A$ , i.e.,  $p(A) = \gcd\{n : \exists q' \in Q', \mathbf{w} \in \Sigma^n \text{ such that } \delta'(q', \mathbf{w}) = q'\}$  and call an automaton with period 1 primitive.

We define for  $M' \subseteq Q'$  and  $\mathbf{w} \in \Sigma^*$

$$\delta'(M', \mathbf{w}) := \{\delta'(q', \mathbf{w}) \mid q' \in M'\}$$

and for a tuple  $(q'_1, \dots, q'_r) \in (Q')^r$

$$\delta'((q'_1, \dots, q'_r), \mathbf{w}) = (\delta'(q'_1, \mathbf{w}), \dots, \delta'(q'_r, \mathbf{w})).$$

We let  $\pi_\ell$  denote the projection on the  $\ell$ -th coordinate (i.e.,  $\pi_\ell(q'_1, \dots, q'_r) = q'_\ell$ ).

**Definition 1.7.** A finite-state transducer is a 6-tupel  $\mathcal{T} = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$ , where  $Q$  is a finite set of states,  $\Sigma$  is the input alphabet,  $\delta$  is the transition function,  $q_0$  is the initial state,  $\Delta$  is the output alphabet and  $\lambda : Q \times \Sigma \rightarrow \Delta^*$  is the output function.

We will restrict ourself to  $\lambda : Q \times \Sigma \rightarrow \Delta$ . Throughout this work we assume that  $(\Delta, \circ)$  is a group generated by  $\text{im}(\lambda)$  and the product in the definition of  $T$  is a product according to  $\circ$ .

A transducer can be viewed as a mean to define functions: on input  $\mathbf{w} = w_1 w_2 \dots w_r \in \Sigma^r$  the transducer enters states  $q_0 = \delta(q_0, \varepsilon), \delta(q_0, w_1), \dots, \delta(q_0, w_1 w_2 \dots w_r)$  and produces the outputs

$$\lambda(q_0, w_1), \lambda(\delta(q_0, w_1), w_2), \dots, \lambda(\delta(q_0, w_1 w_2 \dots w_{r-1}), w_r).$$

The function  $T(\mathbf{w})$  is then defined as

$$T(\mathbf{w}) := \prod_{j=0}^{r-1} \lambda(\delta(q_0, w_1 w_2 \dots w_j), w_{j+1}).$$

We also define the slightly more general form,

$$T(q, \mathbf{w}) := \prod_{j=0}^{r-1} \lambda(\delta(q, w_1 w_2 \dots w_j), w_{j+1}).$$

We define for a set  $M \subseteq \Delta$  the inverse set and the multiplication of two sets  $M_1, M_2 \subseteq \Delta$  as usual,

$$\begin{aligned} M^{-1} &:= \{g^{-1} \mid g \in M\} \\ M_1 \cdot M_2 &:= \{g_1 \circ g_2 \mid g_1 \in M_1, g_2 \in M_2\}. \end{aligned}$$

We say a transducer (or analogously a DFA) is **synchronized** if and only if

$$\exists q \in Q, \mathbf{w}_q \in \Sigma^* \quad \forall q_1 \in Q : \delta(q_1, \mathbf{w}_q) = q.$$

We call  $\mathbf{w}_q$  a **synchronizing word**.

We define for  $\sigma \in S_n$  and a  $n$ -tuple  $\mathbf{x} = (x_1, \dots, x_n)$

$$\sigma \cdot \mathbf{x} := (x_{\sigma^{-1}(1)}, \dots, x_{\sigma^{-1}(n)}),$$

which is the natural way to define the multiplication of a permutation from the left as this ensures associativity:  $(\sigma_1 \circ \sigma_2) \cdot \mathbf{x} = \sigma_1 \cdot (\sigma_2 \cdot \mathbf{x})$  holds for all  $\sigma_1, \sigma_2 \in S_n$ .

Let  $A$  be a  $n \times n$  matrix with complex entries. We let  $A^H$  denote the conjugate transpose of  $A$ . Furthermore, the Frobenius norm of a matrix is given by

$$\|A\|_F := \sqrt{\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2}.$$

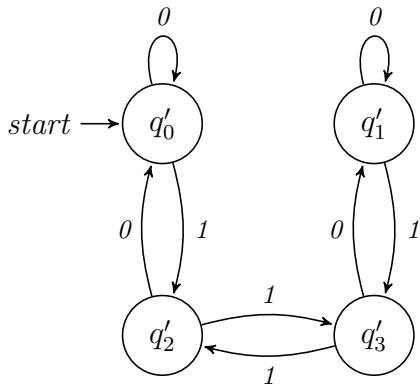
## 2. INDUCED TRANSDUCER

The goal of this section is to define for every strongly connected Automaton  $A = (Q', \Sigma, \delta', q'_0)$  a "naturally induced transducer"  $\mathcal{T}_A = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$  such that we can reconstruct  $\delta'(q'_0, (n)_k)$  by knowing  $T(q_0, (n)_k)$  and  $\delta(q_0, (n)_k)$  for all  $n$ .

The main advantage of this step is that it reveals the actual structure of  $A$ . On a more technical side, we can treat  $\delta$  and  $T$  almost independently which simplifies the analysis of the automatic sequence.

**2.1. Automaton to Transducer.** In this subsection, we develop a special way to construct for an automaton  $A = (Q', \Sigma, \delta', q'_0)$  a "naturally induced" transducer  $\mathcal{T}_A = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$  which contains all the information of  $A$ . The main idea is to merge some states of  $A$  into a single state of  $\mathcal{T}_A$  consistent with  $\delta$  and  $\delta'$ . To illustrate this concept, we start with a simple motivating example:

**Example.** A classical block-additive automatic sequence is the Rudin-Shapiro sequence. The automaton generating this function looks as follows.



One finds easily

$$\delta'(\{q'_0, q'_1\}, 0) = \{q'_0, q'_1\} \quad \delta'(\{q'_0, q'_1\}, 1) = \{q'_2, q'_3\}$$

$$\delta'(\{q'_2, q'_3\}, 0) = \{q'_0, q'_1\} \quad \delta'(\{q'_2, q'_3\}, 1) = \{q'_3, q'_2\}.$$

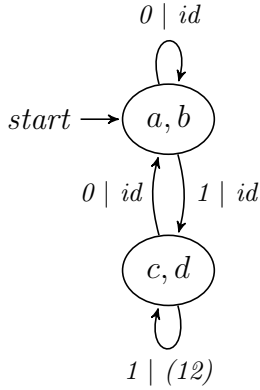
Consequently, we want to merge the pair  $(q'_0, q'_1)$  into a single state  $q_0$  (i.e.,  $q_0 = (q'_0, q'_1)$ ) and  $q_1 = (q'_2, q'_3)$ . We find

$$\begin{aligned} \delta'(q_0, 0) &= q_0 & \delta'(q_0, 1) &= q_1 \\ \delta'(q_1, 0) &= q_0 & \delta'(q_1, 1) &= (12) \cdot q_1. \end{aligned}$$

We define  $(\Delta := S_2)$

$$\begin{aligned} \delta(q_0, 0) &= q_0 & \delta(q_0, 1) &= q_1 \\ \delta(q_1, 0) &= q_0 & \delta(q_1, 1) &= q_1 \\ \lambda(q_0, 0) &= id & \lambda(q_0, 1) &= id \\ \lambda(q_1, 0) &= id & \lambda(q_1, 1) &= (12) \end{aligned}$$

and find that  $\delta'(q_i, \varepsilon) = \lambda(q_i, \varepsilon) \cdot \delta(q_i, \varepsilon)$  holds. This is already the first example of a naturally induced transducer and an illustration of it is given below.



To formalize this idea we first define induced transducers.

**Definition 2.1.** We call a transducer  $\mathcal{T}_A$  **induced** by a DFA  $A = (Q', \Sigma, \delta', q'_0)$  if and only if  $\mathcal{T}_A = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$  where

- 1)  $\exists n_0 \in \mathbb{N} : Q \subseteq (Q')^{n_0}$
- 2)  $\pi_1(q_0) = q'_0$
- 3)  $\Delta < S_{n_0}$  is the subgroup generated by  $Im(\lambda)$
- 4)  $\forall q \in Q, a \in \Sigma : \delta'(q, a) = \lambda(q, a) \cdot \delta(q, a)$ .

We call  $\mathcal{T}_A$  a **naturally induced transducer** if furthermore

- 5)  $\forall i \neq j \leq n_0, q \in Q : \pi_i(q) \neq \pi_j(q)$ ,
- 6)  $\forall q_1 \neq q_2 \in Q \exists \sigma \in S_{n_0} : q_1 = \sigma \cdot q_2$ ,
- 7)  $\mathcal{T}_A$  is strongly connected
- 8)  $\mathcal{T}_A$  is synchronizing

holds.

**Remark.** Properties 5)-7) only assure that  $\mathcal{T}_A$  is chosen minimal to certain aspects.

**Example.** We call  $\mathcal{T}_A$  the *trivial* transducer induced by  $A$  when:

$$n_0 = 1, Q = Q', \Delta = \{id\}, \forall q \in Q, a \in \Sigma : \delta(q, a) = \delta'(q, a), \lambda(q, a) = id.$$

The trivial induced transducer is a naturally induced transducer if and only if  $A$  is synchronizing and strongly connected.

**Proposition 2.2.** For every strongly connected automaton  $A$ , there exists a naturally induced transducer  $\mathcal{T}_A = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$ . All other naturally induced transducers can be obtained by changing the order on the elements of  $Q$  and possibly changing the initial state  $q_0$ .

*Proof.* The proof is not constructive and we start by defining  $n_0(A) := \min\{|\delta'(Q', \mathbf{w})| : \mathbf{w} \in \Sigma^*\}$  and

$$S(A) := \{M \subseteq Q' : |M| = n_0(A), \exists \mathbf{w} \in \Sigma^* \text{ with } \delta'(Q', \mathbf{w}) = M\},$$

which are exactly the sets of size  $n_0(A)$  that are reachable from  $Q'$  using  $\delta'$ . One easily observes that an automaton is synchronizing if and only if  $n_0(A) = 1$  and invertible<sup>4</sup> if and only if  $n_0(A) = |Q'|$ .

Now we are prepared to construct a naturally induced transducer for  $A$ , similar to the previous example.

We choose  $n_0 = n_0(A)$ . Furthermore, we can choose any ordering  $\leq_{Q'}$  of  $Q'$  such that  $q'_0$  is the minimal element. For every  $M \in S(A)$  we define a corresponding  $n_0$ -tuple  $q_M$  which consists of the elements of  $M$  ordered by  $\leq_{Q'}$ . We define  $Q := \{q_M : M \in S(A)\}$ .

Take an arbitrary  $q_M \in Q$ . By minimality of  $|M|$  for  $a \in \Sigma$  we have  $\delta'(M, a) \in S(A)$ . Thus we define  $\delta(q_M, a) := q_{\delta'(M, a)}$ .

We note that  $Q'$  is covered by  $S(A)$ : Take  $q'_1 \in Q'$  and  $M \in S(A)$  with some  $q'_2 \in M$ . As  $A$  is strongly connected, there exists  $\mathbf{w}$  such that  $\delta'(q'_2, \mathbf{w}) = q'_1$ . Thus, we find  $q'_1 \in \delta'(M, \mathbf{w}) \in S(A)$ .

We define  $q_0$  as an arbitrary element  $q_M \in Q$  such that  $q'_0 \in M$ . (Note that  $\pi_1(q_M) = q'_0$  since  $q'_0$  is the minimal element).

We define  $\lambda' : (Q')^{n_0} \rightarrow S_{n_0}$  such that for all  $\tilde{q} = (q'_1, \dots, q'_{n_0}) \in (Q')^{n_0}$ ,  $\lambda'(\tilde{q})$  is the permutation such that  $\lambda'(\tilde{q}) \cdot \tilde{q}$  is ordered with respect to  $\leq_{Q'}$  and  $\lambda : Q \times \Sigma \rightarrow S_{n_0}$ ,  $\lambda(q_M, a) = \lambda'(\delta'(q_M, a))^{-1}$  for all  $M \in S(A)$  and  $a \in \Sigma$ .

Let  $\Delta$  be the subgroup of  $S_{n_0}$  generated by  $Im(\lambda)$ . This defines an induced transducer  $\mathcal{T}_A = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$ .

It remains to show that this induced transducer is a naturally induced transducer. 5) and 6) follow directly from the construction of the transducer. Every  $q_M \in Q$  corresponds to  $M \in S(A)$  for which there exists  $\mathbf{w}_M$  such that  $\delta'(Q', \mathbf{w}_M) = M$ . Thus it follows directly that for all  $q_{M'} \in Q$  it holds  $\delta(q_{M'}, \mathbf{w}_M) = q_M$  which shows that the constructed transducer is synchronizing and also strongly connected.

To prove the second part of this proposition we assume that  $\overline{\mathcal{T}}_A = (\overline{Q}, \Sigma, \overline{\delta}, \overline{q_0}, \overline{\Delta}, \overline{\lambda})$  is an arbitrary naturally induced transducer. We start by showing  $n_0(A) = \overline{n_0}$ .

- Assume  $\overline{n_0} > n_0(A)$ : By the definition of  $n_0(A)$  there exists some  $\mathbf{w} \in \Sigma^*$  such that  $|\delta'(Q', \mathbf{w})| = n_0(A)$ . Consequently  $\overline{\delta}(\overline{q}, \mathbf{w}) \in (Q')^{\overline{n_0}}$  violates 5) for every  $\overline{q} \in \overline{Q}$ .

<sup>4</sup>For the definition of invertible automata see [10].



- Assume  $\bar{n}_0 < n_0(A)$ : As  $\overline{\mathcal{T}_A}$  is synchronizing there exists  $\mathbf{w} \in \Sigma^*$  and  $\bar{q}_1$  such that for all  $\bar{q} \in \overline{Q}$  we have  $\bar{\delta}(\bar{q}, \mathbf{w}) = \bar{q}_1$ . Since  $\overline{Q}$  covers  $Q'$  this implies that  $|\delta'(Q', \mathbf{w})| \leq \bar{n}_0 < n_0(A)$  which contradicts the definition of  $n_0(A)$ .

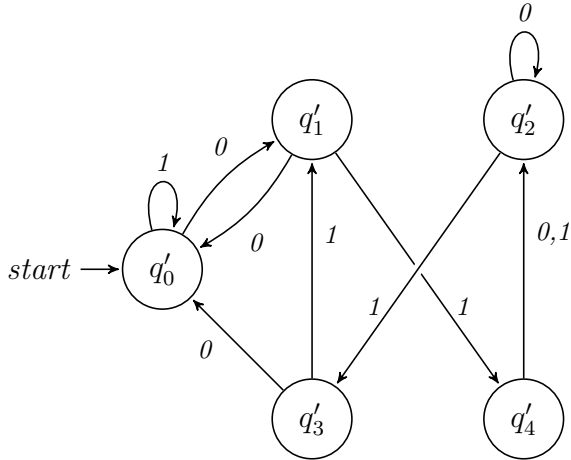
Assume now that there exists some  $q \in Q$  such that no element of  $\overline{Q}$  is a permutation of  $q$ . The elements of  $q$  form a set  $M_q \in S(A)$  and thus there exists  $\mathbf{w}_q$  such that  $\delta'(Q', \mathbf{w}_q) = M_q$ .  $\bar{\delta}(\bar{q}, \mathbf{w}_q)$  contains exactly the elements of  $M_q$  which means that it is a permutation of  $q$  (by property 5)).

The elements of  $\overline{Q}$  that are permutations of elements of  $Q$  form a strongly connected component. Thus, the proposition follows by property 7).  $\square$

**Remark.** Let  $\mathcal{T} = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$  be a transducer fulfilling 1)-5) and 7)-8). We define an equivalence relation  $\sim$  on  $Q$  by:  $q_1 \sim q_2$  if and only if  $\exists \sigma: q_1 = \sigma \cdot q_2$ . Then  $\mathcal{T}/\sim$  is a naturally induced transducer.

**Remark.** One can easily show the minimal length of a synchronizing word for a naturally induced transducer  $\mathcal{T}_A$  is bounded by  $O(|Q'|^3)$ . This already allows to determine  $n_0$ , although in a rather inefficient way, by considering all paths of this length. A more efficient approach would be to determine all sets that can be obtained by  $\delta'(Q', \mathbf{w})$ . This can be done for example by a depth-first search.

**Example.** We consider the following more complex automaton with  $\Sigma = \{0, 1\}$  and find a naturally induced transducer.



One finds that  $n_0(A) = 3$  and  $S(A) = \{M_1, M_2\}$  with

$$M_1 := \delta'(Q', 0) = \{q'_0, q'_1, q'_2\}$$

$$M_2 := \delta'(Q', 01) = \{q'_0, q'_3, q'_4\}.$$

We construct now a naturally induced transducer  $\mathcal{T}_A$  and start by defining  $Q = \{q_0, q_1\}$  where

$$q_0 = (q'_0, q'_1, q'_2)$$

$$q_1 = (q'_0, q'_3, q'_4).$$

We find

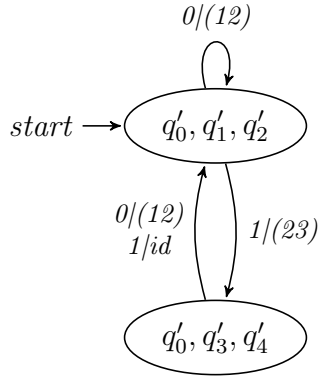
$$\delta'(q_0, 0) = (q'_1, q'_0, q'_2), \quad \delta'(q_0, 1) = (q'_0, q'_4, q'_3)$$

$$\delta'(q_1, 0) = (q'_1, q'_0, q'_2), \quad \delta'(q_1, 1) = (q'_0, q'_1, q'_2)$$

and therefore

$$\begin{aligned} \delta(q_0, 0) &= q_0, & \delta(q_0, 1) &= q_1 \\ \delta(q_1, 0) &= q_0, & \delta(q_1, 1) &= q_0 \\ \lambda(q_0, 0) &= (12), & \lambda(q_0, 1) &= (23) \\ \lambda(q_1, 0) &= (12), & \lambda(q_1, 1) &= id. \end{aligned}$$

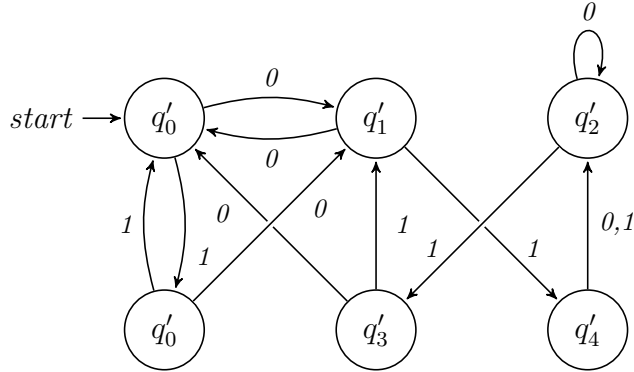
Thus we constructed a transducer  $\mathcal{T}_A$  (as described in the proof of Proposition 2.2):



One easily observes that this is a naturally induced transducer.

**Remark.** It is sometimes useful to consider the automaton where each state occurs as often as it occurs in the subsets of  $S(A)$  and "group" them according to  $S(A)$ .

**Example.** For the example above this gives the following automaton.



We end this subsection with a technical lemma used later.

**Lemma 2.3.** Let  $\mathcal{T}_A$  be a naturally induced transducer. Then there exists  $N_0 \in \mathbb{N}$  such that for all  $q_1, q_2 \in Q$  and  $n \geq N_0$  there exists  $\mathbf{w} \in \Sigma^n$  such that  $\delta(q_1, \mathbf{w}) = q_2$ .

*Proof.* Let  $\mathbf{w}_q$  be a synchronizing word. We denote by  $dist(x, y) := \min\{n \mid \exists \mathbf{w} \in \Sigma^n : \delta(x, \mathbf{w}) = y\}$  and define  $N_0 := |\mathbf{w}_q| + \max_{q' \in Q} dist(q, q')$ .

Take  $q_1, q_2 \in Q$  and  $n \geq N_0$ . There exists  $\mathbf{w}_1$  from  $q$  to  $q_2$  of length  $dist(q, q_2)$ . We consider  $\mathbf{w} = 0^{n-|\mathbf{w}_q|-dist(q, q_2)} \mathbf{w}_q \mathbf{w}_1$  and see that this is a path from  $q_1$  to  $q_2$  of length  $n$  ( $0^x$  denotes the word consisting of  $x$  consecutive zeros).  $\square$

**2.2. Connection between an automaton and its induced transducer.** Property 4) of an induced transducer relates one step of  $A$  to one step of  $\mathcal{T}_A$  and  $\lambda/T$ . In this section we show how arbitrary many steps of  $A$  relate to  $\mathcal{T}_A$ . Therefore we start with a technical lemma that describes how the action of a permutation to a state  $q \in Q$  influences the behavior of the induced transducer.

**Lemma 2.4.** *Let  $A$  be a DFA and  $\mathcal{T}_A$  an induced transducer. Furthermore, let  $\sigma \in S_{n_0}$ ,  $a \in \Sigma$  and  $q = (x_1, \dots, x_{n_0}) \in Q$ . It holds*

$$\delta'(\sigma \cdot q, a) = (\sigma \circ \lambda(q, a)) \cdot \delta(q, a).$$

*Proof.* We find for  $1 \leq k \leq n_0$

$$\begin{aligned} \pi_k(\delta'(\sigma \cdot (x_1, \dots, x_{n_0}), a)) &= \delta'(x_{\sigma^{-1}(k)}, a) = \pi_{\sigma^{-1}(k)}(\delta'(q, a)) \\ &= \pi_{\sigma^{-1}(k)}(\lambda(q, a) \cdot \delta(q, a)) = \pi_k(\sigma \cdot (\lambda(q, a) \cdot \delta(q, a))). \end{aligned}$$

Thus, it holds

$$\delta'(\sigma \cdot q, a) = \sigma \cdot (\lambda(q, a) \cdot \delta(q, a)) = (\sigma \circ \lambda(q, a)) \cdot \delta(q, a).$$

□

Now we are ready to show a very important connection between an automaton and its induced transducer.

**Proposition 2.5.** *Let  $A$  be a strongly connected automaton and  $\mathcal{T}_A$  an induced transducer. For every  $\mathbf{w} \in \Sigma^*$  we have*

$$\delta'(q'_0, \mathbf{w}) = \pi_1(T(q_0, \mathbf{w}) \cdot \delta(q_0, \mathbf{w})).$$

*Proof.* We actually prove that for all  $q \in Q$  and  $\mathbf{w} \in \Sigma^*$  it holds that

$$\delta'(q, \mathbf{w}) = T(q, \mathbf{w}) \cdot \delta(q, \mathbf{w}),$$

which obviously implies the statement. We use induction on the length of  $\mathbf{w}$ .

- $|\mathbf{w}| = 0$ : Obviously  $\delta'(q, \varepsilon) = q = (id(\delta(q, \varepsilon)))$  holds.
- $\mathbf{w} = \mathbf{w}'x$ : We define  $q_1 := \delta(q, \mathbf{w}')$  and find

$$\begin{aligned} \delta'(q, \mathbf{w}'x) &= \delta'(\delta'(q, \mathbf{w}'), x) = \delta'(T(q, \mathbf{w}') \cdot q_1, x) = (T(q, \mathbf{w}') \circ \lambda(q_1, x)) \cdot \delta(q_1, x) \\ &= T(q, \mathbf{w}'x) \cdot \delta(q, \mathbf{w}'x), \end{aligned}$$

where the third equality holds by Lemma 2.4.

□

This Proposition allows us to reconstruct the output of  $A$  by knowing the output of  $\mathcal{T}_A$  and  $T$ .

**Example.** *We continue our previous example and find*

$$\begin{aligned} \delta'(q'_0, 0110) &= q'_2 = \pi_1(T(q_0, 0110) \cdot \delta(q_0, 0110)) = \pi_1(((12) \circ (23) \circ id \circ (12)) \cdot q_0) \\ &= \pi_1((13) \cdot q_0) = \pi_1(q'_2, q'_1, q'_0) = q'_2. \end{aligned}$$

As one has some freedom on how to choose the order of the elements of  $Q$  it is natural to ask how this choice influences the induced transducer.

**Proposition 2.6.** *Let  $A$  be a DFA and  $\mathcal{T}_A = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$  an induced transducer. Let  $\overline{\mathcal{T}}_A$  be another induced transducer obtained by changing the order of every tuple  $q \in Q$  by a permutation  $\sigma_q$ , i.e.,  $\overline{q} = \sigma_q \cdot q$  (where still  $\pi_1(\sigma_{q_0} \cdot q_0) = q'_0$  holds). For  $\mathbf{w} \in \Sigma^*$  it holds that*

$$\overline{T}(\overline{q}, \mathbf{w}) = \sigma_q \circ T(q, a) \circ \sigma_{\delta(q, \mathbf{w})}^{-1}.$$

Furthermore, if  $\mathcal{T}_A$  is a naturally induced transducer, then so is  $\overline{\mathcal{T}}_A$ .

*Proof.* Let  $q \in Q, a \in \Sigma$ . We define for all  $q \in Q, a \in \Sigma$

$$\begin{aligned} \overline{q} &:= \sigma_q \cdot q \\ \overline{\delta}(\overline{q}, a) &:= \overline{\delta(q, a)} \\ \overline{\lambda}(\overline{q}, a) &:= \sigma_q \circ \lambda(q, a) \circ \sigma_{\delta(q, a)}^{-1}, \end{aligned}$$

$\overline{Q} := \cup_{q \in Q} \overline{q}$  and  $\overline{\Delta}$  is again the group generated by  $\overline{\lambda}$ .

We claim that  $\overline{\mathcal{T}}_A = (\overline{Q}, \Sigma, \overline{\delta}, \overline{q}_0, \overline{\Delta}, \overline{\lambda})$  is again an induced transducer. 1)-3) follow immediately so it remains to show 4). Therefore, we compute by Lemma 2.4

$$\begin{aligned} \delta'(\sigma_q \cdot q, a) &= (\sigma_q \circ \lambda(q, a)) \cdot \delta(q, a) \\ &= (\sigma_q \circ \lambda(q, a) \circ \sigma_{\delta(q, a)}^{-1}) \cdot (\sigma_{\delta(q, a)} \cdot \delta(q, a)) \\ &= \overline{\lambda}(\overline{q}, a) \cdot \overline{\delta}(\overline{q}, a). \end{aligned}$$

Thus  $\overline{\mathcal{T}}_A$  is indeed an induced transducer and the stated equation follows easily by induction on the length of  $\mathbf{w}$ .

The last statement is very easy to verify. □

**2.3. Length restrictions for naturally induced transducers.** As  $\Delta$  is generated by  $\lambda$ , one may assume that all elements of  $\Delta$  can be obtained by  $T$ . We show in this subsection that this is not true in general but for certain naturally induced transducers. Furthermore, we will show that restrictions on the length of  $\mathbf{w}$  give restrictions on what elements of  $\Delta$  can be obtained by  $T(q, \mathbf{w})$ .

The main result of this subsection – which contains all the facts mentioned above – is the following theorem.

**Theorem 2.7.** *Let  $A$  be a strongly connected automaton. Then there exists a minimal  $d \in \mathbb{N}$ ,  $m_0 \in \mathbb{N}$ , a naturally induced transducer  $\mathcal{T}_A$  and a subgroup  $G$  of  $\Delta$  such that the following two conditions hold.*

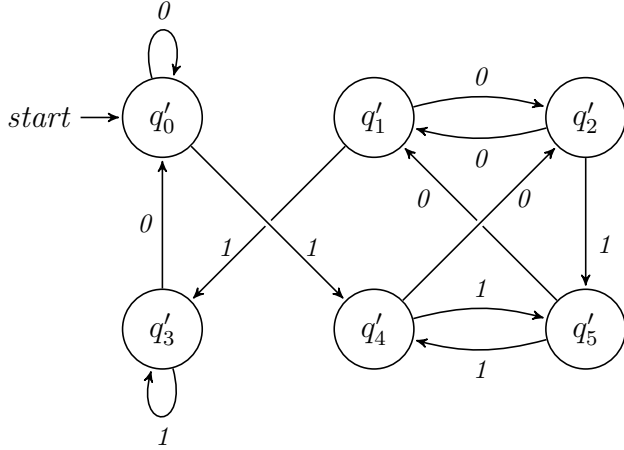
- For all  $q \in Q, \mathbf{w} \in (\Sigma^d)^*$  we have  $T(q, \mathbf{w}) \in G$ .
- For all  $q, \overline{q} \in Q$  and  $n \geq m_0$  it holds that

$$\{T(q, \mathbf{w}) : \mathbf{w} \in \Sigma^{nd}, \delta(q, \mathbf{w}) = \overline{q}\} = G.$$

$d$  and  $m_0$  only depend on  $A$ , but not on its initial state  $q'_0$ .

We provide an example of a primitive (and strongly connected) automaton for which  $d > 1$  holds.

**Example.** Consider the following automaton.



This automaton is clearly primitive, i.e., its period is 1, as  $\delta(q_0, 0) = q_0$ . One choice for a naturally induced transducer is already indicated by the sketch above,

$$\begin{aligned} q_0 &= (q'_0, q'_1, q'_2), & q_1 &= (q'_3, q'_4, q'_5) \\ \delta(q_0, 0) &= q_0, & \delta(q_0, 1) &= q_1 \\ \delta(q_1, 0) &= q_0, & \delta(q_1, 1) &= q_1 \\ \lambda(q_0, 0) &= (23), & \lambda(q_0, 1) &= (12) \\ \lambda(q_1, 0) &= (23), & \lambda(q_1, 1) &= (23). \end{aligned}$$

One sees directly that  $\text{sgn}(\lambda(q, a)) = -1$  for all  $a \in \Sigma$  and therefore

$$T(q, \mathbf{w}) \in \begin{cases} \{id, (123), (132)\}, & \text{for } \mathbf{w} \in (\Sigma^2)^* \\ \{(12), (23), (13)\}, & \text{for } \mathbf{w} \in \Sigma(\Sigma^2)^* \end{cases}$$

Indeed, one can show that  $d = 2$ ,  $G = \{id, (123), (132)\}$  and  $g_0 = (12)$ , as all elements of  $G$  appear as output for paths from  $q_0$  to itself for length 4 as well as for length 6.

The other results of this subsection are rather technical; they will be used only in this and the following subsection.

According to Proposition 2.2 we assume in this section that  $\mathcal{T}_A$  is a naturally induced transducer of a strongly connected automaton. Thus,  $\Delta$  is finite.

Paths  $\mathbf{w}$  with the property

$$(1) \quad \delta(q, \mathbf{w}) = q, T(q, \mathbf{w}) = id$$

will play an important role in this subsection and we start by exploring some of their properties.

**Remark.** Obviously concatenating two paths  $\mathbf{w}_1, \mathbf{w}_2$  fulfilling (1) gives again a path fulfilling (1).

We start this section by showing that every path can be extended to a path with this property.

**Lemma 2.8.** For every  $q \in Q$ ,  $\mathbf{w} \in \Sigma^*$  there exists an *inverse path*  $\overline{\mathbf{w}}^q \in \Sigma^*$  such that (1) holds for  $\mathbf{w}\overline{\mathbf{w}}^q$ .

*Proof.* Since  $\mathcal{T}_A$  is strongly connected, there exists  $\mathbf{w}' \in \Sigma^*$  such that  $\delta(q, \mathbf{w}\mathbf{w}') = \delta(\delta(q, \mathbf{w}), \mathbf{w}') = q$ .

We define  $g := T(q, \mathbf{w}\mathbf{w}')$ . Since  $G$  is finite we know that  $g^n = id$  for some  $n \in \mathbb{N}$ . Since  $\delta(q, (\mathbf{w}\mathbf{w}')^n) = q$  and  $T(q, (\mathbf{w}\mathbf{w}')^n) = id$  we know that the desired property holds for  $\overline{\mathbf{w}}^q := \mathbf{w}'(\mathbf{w}\mathbf{w}')^{n-1}$ .  $\square$

We define  $M_q := \{n \in \mathbb{N} \mid \exists \mathbf{w} \in \Sigma^n, \delta(q, \mathbf{w}) = q, T(q, \mathbf{w}) = id\}$  and it follows directly by Lemma 2.8 that  $M_q$  is non-empty for all  $q$ .

**Lemma 2.9.** *Let  $A$  be a strongly connected automaton and  $\mathcal{T}_A$  an induced transducer. There exists  $d = d(\mathcal{T}_A) \in \mathbb{N}$  such that for all  $q \in Q$  :  $M_q \subset d \cdot \mathbb{N}$  and  $d \cdot \mathbb{N} \setminus M_q$  is finite.*

*Proof.* The first remark in this subsection directly shows that  $M_q$  is closed under addition for all  $q \in Q$ . Thus, it is easy to show that for every  $q \in Q$  there exists  $d_q$  such that  $M_q \subset d_q \cdot \mathbb{N}$  and  $d_q \cdot \mathbb{N} \setminus M_q$  is finite:

We take  $d_q := \gcd(M_q)$  which can be written as  $d_q = \sum_{\ell < r} \alpha_\ell n_\ell$  for some  $r \in \mathbb{N}$  and  $n_\ell \in M_q$ . Obviously, we find  $M_q \subset d_q \cdot \mathbb{N}$ . Let  $m := \sum_{\ell < r} n_\ell$ , then we can write any  $N \in d_q \cdot \mathbb{N}$  as  $rm + sd_q$  with  $0 \leq s < m/d_q$  and one finds for sufficiently large  $N$  - and, therefore, sufficiently large  $r$  - that all coefficients of the  $n_\ell$  are positive and, therefore,  $N \in M_q$ .

It remains to show that all  $d_q$  coincide. Let  $q_1, q_2 \in Q$ . There exist  $m \in \mathbb{N}$ ,  $\mathbf{w}_1 \in \Sigma^m$  and  $\mathbf{w}_2 \in \Sigma^{m+d_{q_1}}$  such that (1) holds for  $\mathbf{w}_1, \mathbf{w}_2$  (for  $q = q_1$ ). Furthermore, by Lemma 2.3 and Lemma 2.8, there exist  $\mathbf{w}, \overline{\mathbf{w}}^{q_2}$  such that  $\delta(q_2, \mathbf{w}) = q_1, \delta(q_1, \overline{\mathbf{w}}^{q_2}) = q_2$  and  $T(q_2, \mathbf{w}\overline{\mathbf{w}}^{q_2}) = id$ . Therefore,  $\mathbf{w}\mathbf{w}_i\overline{\mathbf{w}}^{q_2}$  fulfill (1) for  $i = 1, 2$  (and  $q = q_2$ ). These two paths belonging to  $M_{q_2}$  have length  $|\mathbf{w}| + |\overline{\mathbf{w}}^{q_2}| + m$  and  $|\mathbf{w}| + |\overline{\mathbf{w}}^{q_2}| + m + d_{q_1}$  respectively. As  $M_{q_2} \subseteq d_{q_2}\mathbb{N}$  we see that  $d_{q_2} \leq d_{q_1}$  holds. Since  $q_1, q_2$  are arbitrary elements of  $Q$  the statement follows.  $\square$

**Example.** *Consider the following naturally induced transducer with one state, where  $k = 2$ .*

$$\lambda(q_0, 0) = (23)$$

$$\lambda(q_0, 1) = (123)$$

*One finds easily that  $d = 1$  and  $M_{q_0} = \mathbb{N} \setminus \{1\}$ .*

**Remark.** *Changing the order of the tuples  $q \in Q$ , i.e., considering a different naturally induced transducer, does not change  $d(\mathcal{T}_A)$  (since  $1 \mapsto \sigma_q \circ id \circ \sigma_q^{-1} = id$ ). Thus we find that  $d(\mathcal{T}_A)$  only depends on  $A$  and we write  $d = d(A) = d(\mathcal{T}_A)$ .*

As mentioned before, we are interested in which elements of  $\Delta$  occur for paths with certain length. One might assume that some periodic behavior (with period  $d$ ) might occur when restricting to paths of certain length and, furthermore, that for long enough paths some closure properties hold.

To find some precise statements, we define

$$M_{q\bar{q}, \ell} := \{T(q, \mathbf{w}) \mid \mathbf{w} \in \Sigma^\ell, \delta(q, \mathbf{w}) = \bar{q}\},$$

for non-negative  $\ell$  and  $M_{q\bar{q}, \ell} = \emptyset$  for negative  $\ell$ .

$M_{q\bar{q}, \ell}$  are all the group elements that occur when reading a word of length  $\ell$  that corresponds to a path from  $q$  to  $\bar{q}$ . Furthermore, we define  $G_{q\bar{q}}(\ell) := \bigcup_{\ell' \equiv \ell (d)} M_{q\bar{q}, \ell'}$ . The following lemma shows that this union is actually a limit.

**Lemma 2.10.** *For all  $q, \bar{q} \in Q$  and for all  $\ell$  such that  $0 \leq \ell \leq d-1$  it holds that,  $G_{q\bar{q}}(\ell) = M_{q\bar{q}, \ell+kd}$  for almost all - i.e., all but finitely many -  $k \in \mathbb{N}$ .*

*Proof.* It is sufficient to show that for each  $g \in G_{q\bar{q}}(\ell)$  we have  $g \in M_{q\bar{q}, \ell+kd}$  for almost all  $k \in \mathbb{N}$ .

For  $g \in G_{q\bar{q}}(\ell)$ , there exists  $k_0 \in \mathbb{N}$  and  $\mathbf{w} \in \Sigma^{\ell+k_0d}$  such that  $\delta(q, \mathbf{w}) = \bar{q}$  and  $T(q, \mathbf{w}) = g$ . Lemma 2.9 implies that we have for almost all  $k' \in \mathbb{N}$ : There exists  $\mathbf{w}_{k'} \in \Sigma^{k'd}$  which fulfills (1). Since  $\mathbf{w}_{k'}\mathbf{w}$  is a path from  $q$  to  $\bar{q}$  with weight  $g$ , we have for almost all  $k \geq k_0$  (and therefore for almost all  $k \in \mathbb{N}$ ),  $g \in M_{q\bar{q}, \ell+kd}$ .  $\square$

We are interested in how the different  $G_{q\bar{q}}(\ell)$  are related. We see easily that for  $q_1, q_2, q_3 \in Q$ ,  $G_{q_1q_2}(\ell_1) \cdot G_{q_2q_3}(\ell_2) \subseteq G_{q_1q_3}(\ell_1 + \ell_2)$  holds, but we actually even find equality:

**Lemma 2.11.** *For all  $\ell_1, \ell_2 \in \mathbb{N}$  and  $q_1, q_2, q_3 \in Q$  it holds that*

$$G_{q_1q_3}(\ell_1 + \ell_2) = G_{q_1q_2}(\ell_1) \cdot G_{q_2q_3}(\ell_2).$$

*Furthermore,  $|G_{q_0q_0}(0)| = |G_{q\bar{q}}(\ell)|$  holds for all  $q, \bar{q} \in Q$  and  $\ell$  such that  $0 \leq \ell \leq d-1$ .*

*Proof.* We start by showing  $G_{q\bar{q}}(\ell)^{-1} = G_{\bar{q}q}(-\ell)$  for all  $q, \bar{q} \in Q, \ell \in \mathbb{N}$ .

We know by Lemma 2.8 that for each  $\mathbf{w}$  with  $\delta(q, \mathbf{w}) = \bar{q}$ , there exists  $\bar{\mathbf{w}}^q$ . Thus we conclude that  $|\mathbf{w}| + |\bar{\mathbf{w}}^q| \equiv 0 \pmod{d}$  and, therefore,  $G_{q\bar{q}}(\ell)^{-1} \subseteq G_{\bar{q}q}(-\ell)$ . By inverting both sides, we also find  $G_{q\bar{q}}(\ell) \subseteq G_{\bar{q}q}(-\ell)^{-1}$  and since  $q, \bar{q}$  and  $\ell$  are chosen arbitrarily,  $G_{q\bar{q}}(\ell)^{-1} = G_{\bar{q}q}(-\ell)$  holds for all  $q, \bar{q} \in Q, \ell \in \mathbb{N}$ .

The inclusion  $\supseteq$  is trivial. Let  $g \in G_{q_1q_3}(\ell_1 + \ell_2)$ . By definition there exists  $\mathbf{w}$  such that  $\delta(q_1, \mathbf{w}) = q_3, T(q_1, \mathbf{w}) = g$  and  $|\mathbf{w}| \equiv \ell_1 + \ell_2 \pmod{d}$ . Lemma 2.3 shows the existence of  $\mathbf{w}_1$  with  $\delta(q_1, \mathbf{w}_1) = q_2$  and  $|\mathbf{w}_1| \equiv \ell_1 \pmod{d}$ . We find, by the first part of this proof, an inverse path  $\bar{\mathbf{w}}_1^{q_1}$  from  $q_2$  to  $q_1$  such that  $|\bar{\mathbf{w}}_1^{q_1}| \equiv -\ell_1 \pmod{d}$  holds. We see that  $\mathbf{w}_2 := \bar{\mathbf{w}}_1^{q_1}\mathbf{w}$  is a path from  $q_2$  to  $q_3$  with  $|\mathbf{w}_2| \equiv \ell_2 \pmod{d}$  and

$$\begin{aligned} T(q_1, \mathbf{w}_1\mathbf{w}_2) &= T(q_2, \mathbf{w}_1) \cdot T(q_1, \mathbf{w}_2) = T(q_1, \mathbf{w}_1) \cdot T(q_2, \bar{\mathbf{w}}_1^{q_1}) \cdot T(q_1, \mathbf{w}) \\ &= T(q_1, \mathbf{w}_1\bar{\mathbf{w}}_1^{q_1}) \cdot g = g. \end{aligned}$$

This finishes the proof of the first statement. We see by the first part of this lemma, that  $G_{q_1q_3}(\ell) \supseteq G_{q_1q_2}(\ell') \cdot g_2$  and  $G_{q_1q_3}(\ell) \supseteq g_1 \cdot G_{q_2q_3}(\ell')$  for  $g_1 \in G_{q_1q_2}(\ell - \ell'), g_2 \in G_{q_2q_3}(\ell - \ell')$ . Thus we find

$$\begin{aligned} |G_{q_1q_3}(\ell)| &\geq |G_{q_1q_2}(\ell')| \\ |G_{q_1q_3}(\ell)| &\geq |G_{q_2q_3}(\ell')| \end{aligned}$$

for all  $q_1, q_2, q_3 \in Q$  and  $\ell, \ell' \in \mathbb{N}$ . We can use this fact to show that

$$|G_{q_0q_0}(0)| \geq |G_{q_0q}(0)| \geq |G_{qq}(\ell)| \geq |G_{q_0q}(0)| \geq |G_{q_0q_0}(0)|.$$

Therefore we see that  $|G_{qq}(\ell)| = |G_{q_0q_0}(0)|$  for all  $q \in Q$  and  $\ell \in \mathbb{N}$ . To complete the proof we see that

$$|G_{q_0q_0}(0)| = |G_{qq}(0)| \geq |G_{q\bar{q}}(\ell)| \geq |G_{\bar{q}q}(0)| = |G_{q_0q_0}(0)|.$$

$\square$

**Remark.** Lemma 2.11 gives a strong structural results. Take, for example, any element  $g_2 \in G_{q_2q_3}(\ell_2)$ . As  $|G_{q_1q_3}(\ell_1 + \ell_2)| = |G_{q_1q_2}(\ell_1)|$ , it follows that  $G_{q_1q_3}(\ell_1 + \ell_2) = G_{q_1q_2}(\ell_1) \cdot g_2$ . A similar result holds for  $g_1 \in G_{q_1q_2}(\ell_1)$ .

**Corollary 2.12.**  $G_{qq}(0)$  is a subgroup of  $\Delta$  for all  $q \in Q$ .

$G_{qq}(0)$  being a subgroup is the first example for the appearance of some closure property occurring in this setting.

In this subsection, we also prove that all  $G_{q\bar{q}}(\ell)$  are cosets of  $G_{q_0q_0}(0)$ , but first we mention a slightly stronger result than Lemma 2.10 which we need in the next subsection.

**Lemma 2.13.** *There exists  $m_0 = m_0(A)$  such that for all  $q, \bar{q} \in Q$  and for all  $\ell$  such that  $0 \leq \ell \leq d - 1, g \in G_{q\bar{q}}(\ell)$  and  $k \geq m_0$  there exist  $\mathbf{w}_1, \mathbf{w}_2 \in \Sigma^{\ell+k d}$  such that  $\mathbf{w}_1 \neq \mathbf{w}_2$  and  $\delta(q, \mathbf{w}_i) = \bar{q}, T(q, \mathbf{w}_i) = g$ .*

*Proof.* Lemma 2.10 guarantees the existence of some  $m_1$  such that for all  $q, \bar{q} \in Q, 0 \leq \ell \leq d - 1, g \in G_{q\bar{q}}(\ell)$  and  $k' \geq m_1$  exists  $\mathbf{w} \in \Sigma^{\ell+k'd}$  such that  $\delta(q, \mathbf{w}) = \bar{q}, T(q, \mathbf{w}) = g$ . We will show that the desired property holds for  $k = 2m_1$ , which implies the statement, since there exists for all  $k' \geq m_1$  a path  $\mathbf{w} \in \Sigma^{k'd}$  with  $\delta(\bar{q}, \mathbf{w}) = \bar{q}$  and  $T(\bar{q}, \mathbf{w}) = id$ .

- Case  $|Q| \geq 2$ :

We fix  $q_1 \neq q_2 \in Q$ . Lemma 2.3 yields the existence of some paths  $\mathbf{w}'_1, \mathbf{w}'_2$  of length  $m_1 d$  such that  $\delta(q, \mathbf{w}'_i) = q_i$ . Of course, we find  $T(q, \mathbf{w}'_i) \in G_{qq_i}(0)$  which shows together with Lemma 2.11 that  $T(q, \mathbf{w}'_i)^{-1}g \in G_{q_i\bar{q}}(\ell)$ . Lemma 2.10 shows that there exist paths  $\mathbf{w}''_1, \mathbf{w}''_2$  of length  $\ell + m_1 d$  such that  $\delta(q_i, \mathbf{w}''_i) = \bar{q}$  and  $T(q, \mathbf{w}'_i \mathbf{w}''_i) = g$ .

- Case  $|Q| = 1, |G_{q_0q_0}(0)| \geq 2$ :

Thus we have  $Q = \{q_0\}$ . We fix  $g_1 \neq g_2 \in G_{q_0q_0}(0)$ . A combination of Lemma 2.10 and Lemma 2.11 yields the existence of some paths  $\mathbf{w}'_1, \mathbf{w}'_2$  of length  $m_1 d$  and paths  $\mathbf{w}''_1, \mathbf{w}''_2$  of length  $\ell + m_1 d$  such that  $T(q_0, \mathbf{w}'_i) = g_i$  and  $T(q_0, \mathbf{w}'_i \mathbf{w}''_i) = g$ .

- Case  $|Q| = |G_{q_0q_0}(0)| = 1$ :

Obviously taking any two different  $\mathbf{w}'_i \in \Sigma^{m_1 d}$  and  $\mathbf{w} \in \Sigma^{\ell+m_1 d}$  such that  $T(q_0, \mathbf{w}'_i) = id, T(q_0, \mathbf{w}) = g$ , gives  $T(q_0, \mathbf{w}'_i \mathbf{w}) = g$ .

Thus we have constructed in all cases two different paths with the desired properties.  $\square$

We have seen in Proposition 2.6 that permuting the elements of  $Q$  again gives an induced transducer. We now show that by choosing the right permutations, we are able to make the actual structure of the induced transducer more apparent.

**Lemma 2.14.** *Let  $\mathcal{T}_A$  be a naturally induced transducer. There exists a naturally induced transducer  $\overline{\mathcal{T}}_A$  such that  $id \in \overline{G}_{q_0\bar{q}}(0)$  holds for each  $\bar{q} \in \overline{Q}$ .*

*Proof.* The idea is to find permutations  $\sigma_q$  such that the induced transducer defined in Proposition 2.6 has the desired properties. We take  $\sigma_q \in G_{q_0q}(0)$  for all  $q \neq q_0$  and  $\sigma_{q_0} = id$ . By Proposition 2.6, we find  $\overline{G}_{q_0\bar{q}}(0) = G_{q_0q} \cdot \sigma_q^{-1}$ , which directly finishes the proof.  $\square$

We consider, from now on, only naturally induced transducers for which  $id \in G_{q_0q}(0)$  holds for all  $q \in Q$ .



We define  $G := G_{q_0q_0}(0)$  and find that the condition above is already sufficient to make  $G_{q\bar{q}}(\ell)$  independent of  $q, \bar{q}$ , which makes the actual structure more apparent.

**Proposition 2.15.** *There exists  $g_0 \in \Delta$  such that we have for all  $q, \bar{q} \in Q, \ell \in \mathbb{N}$*

$$G_{q\bar{q}}(\ell) = G \cdot g_0^\ell = g_0^\ell \cdot G.$$

**Remark.** *It easily follows that  $d | \text{ord}(g_0)$  and for  $d > 1, g_0 \notin G$ .*

*We also find that  $\Delta = G \cdot \{g_0^\ell : \ell \in \mathbb{N}\} = \{g_0^\ell : \ell \in \mathbb{N}\} \cdot G$ , i.e.,  $G \triangleleft \Delta$ .*

*This proposition shows together with Lemma 2.10 that  $M_{q\bar{q}, \ell}$  depends only on  $(\ell \bmod d)$  for large  $\ell$ .*

*Proof.* As a first step, we prove that for all  $q, \bar{q} \in Q$  we have  $G_{q\bar{q}}(0) = G$ :

We know by Lemma 2.11 that  $G_{q_0q}(0) = G_{q_0q_0}(0) \cdot G_{q_0q}(0)$  and without loss of generality by Lemma 2.14,  $id \in G_{q_0q_0}(0)$ . Thus we have  $G_{q_0q_0}(0) \cdot G_{q_0q}(0) \supseteq G \cdot id$  and, by comparing the cardinalities, we see that  $G_{q_0q}(0) = G$ . Analogously, one finds that  $G_{q_0\bar{q}}(0) = G$ . Lemma 2.11 gives again  $G_{q_0\bar{q}}(0) = G_{q_0q}(0) \cdot G_{q\bar{q}}(0)$  (i.e.,  $G = G \cdot G_{q\bar{q}}(0)$ ) and one concludes as above that  $G_{q\bar{q}}(0) = G$ .

Now we consider  $G_{q_0q_0}(1)$ . We find by Lemma 2.11

$$G_{q_0q_0}(1) = G \cdot G_{q_0q_0}(1) = G_{q_0q_0}(1) \cdot G.$$

We now select an arbitrary element  $g_0$  of  $G_{q_0q_0}(1)$  and find

$$G_{q_0q_0}(1) = G \cdot g_0 = g_0 \cdot G.$$

One easily shows by induction that  $G_{q_0q_0}(\ell) = G \cdot g_0^\ell = g_0^\ell \cdot G$  holds for all  $\ell \in \mathbb{N}$ .

It just remains to note that

$$G_{q\bar{q}}(\ell) = G_{q_0q_0}(0) \cdot G_{q_0q_0}(\ell) \cdot G_{q_0\bar{q}}(0) = G \cdot (g_0^\ell \cdot G) \cdot G = g_0^\ell \cdot G.$$

□

Now we can prove Theorem 2.7.

*Proof of Theorem 2.7:* We choose all the variables as defined throughout this section. The statement of this theorem easily follows by a combination of Lemma 2.10 and Proposition 2.15. □

**Remark.** *For the period  $p$  of  $A$ , one easily finds that  $p | d$ . However,  $p = d$  need not hold as the example provided after Theorem 2.7 shows.*

**2.4. Arithmetic restriction for naturally induced transducers.** Theorem 2.7 shows that all elements of  $G$  occur when we restrict ourself to paths whose length is divisible by  $d$ . We show in this subsection that – similarly to Section 2.3 – restrictions on  $[\mathbf{w}]_k \bmod (k^d - 1)$  lead to restrictions on what elements of  $G$  occur.

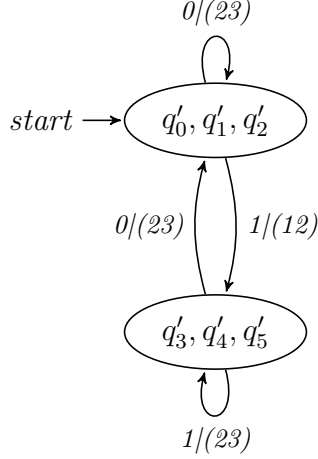
The main result of this subsection is the following theorem.

**Theorem 2.16.** *Under the same conditions as in Theorem 2.7, there exist natural numbers  $k_0, m'_0, \ell_0$  and  $d'$  (where  $d' | k^d - 1$  and therefore  $\gcd(d', k) = 1$ ) together with a naturally induced transducer  $\mathcal{T}_A$  fulfilling the properties of Theorem 2.7, a subgroup  $G_0$  of  $G$  and  $g'_0 \in G$  fulfilling the following two conditions.*

- For all  $q, \bar{q} \in Q$  it holds that
 
$$\{T(q, \mathbf{w}) : q \in Q, \mathbf{w} \in (\Sigma^{dk_0})^*, \delta(q, \mathbf{w}) = \bar{q}, [\mathbf{w}]_k \equiv \ell \pmod{d'}\} = G_0 \cdot g_0^\ell = g_0^{\ell'} \cdot G_0.$$
- For all  $q, \bar{q} \in Q$  there exists  $d''(q, \bar{q}) | k^{\ell_0}$  such that for all  $g \in G_0$  and  $m \geq m'_0$  holds
 
$$\gcd\{[\mathbf{w}]_k : \mathbf{w} \in \Sigma^{dk_0 m}, \delta(q, \mathbf{w}) = \bar{q}, T(q, \mathbf{w}) = g\} = d' \cdot d''(q, \bar{q}).$$

All of the variables only depend on  $A$ , but not on its initial state  $q'_0$ .

**Example.** We continue the example from Section 2.3. The naturally induced transducer we considered was



We have already seen that  $d = 2$  holds. First we would like to mention that it is not necessary to know the values of  $d'$  and  $d''$  to construct the naturally induced transducer, as the proof of Lemma 2.22 shows. Now we derive  $d'$  and  $d''$ . We find easily that  $\mathbf{w}_{q_0} = 0, \mathbf{w}_{q_1} = 1$  are synchronizing words. Furthermore, a path ends in  $q_0$  if and only if it ends with 0 and a path ends in  $q_1$  if and only if it ends with 1. As we will see,  $d''(q, \bar{q})$  corresponds to arithmetic restrictions for paths from  $q$  to  $\bar{q}$ . Thus, we find  $d''(\cdot, q_0) = d''(\cdot, q_1) = 2$ . In general it is sufficient to consider all paths of length at most  $|Q| - 1$  to determine  $d''$ .

We compute

$$\begin{aligned} \delta(q_1, 00) &= \delta(q_1, 10) = q_0 \\ T(q_1, 00) &= T(q_1, 10) = id. \end{aligned}$$

Thus we find  $2 | d(q_1, q_0)$ ,  $d' = 1$  and  $k_0 = 1$ .

In general it is more involved to determine  $d'$ . One straightforward way is to determine  $d_{id, \ell}^{q_0 q_0}$  (defined shortly) for increasing values of  $\ell$  until the conditions of Theorem 2.16 are fulfilled for  $d' \cdot d''(q_0, q_0) = d_{id, \ell}^{q_0 q_0}$ .

The rest of this subsection is again rather technical and only used to prove this theorem.

We only consider  $G$  (i.e., weights corresponding to paths whose length is divisible by  $d$ ) as just the same phenomena occur in the general situation. To assure that we are only working on these elements, we only consider paths whose length is divisible by  $d$ .

Lemma 2.13 assures, that we have for  $\ell \geq m_0$

$$\forall q, \bar{q} \in Q \quad \forall g \in G \quad \exists \mathbf{w}_1, \mathbf{w}_2 \in \Sigma^{d\ell} : \mathbf{w}_1 \neq \mathbf{w}_2, \delta(q, \mathbf{w}_i) = \bar{q}, T(q, \mathbf{w}_i) = g.$$

Therefore, we define for  $\ell \geq m_0, g \in G$

$$d_{g,\ell}^{q\bar{q}} := \max\{m \in \mathbb{N} \mid \exists r : \forall \mathbf{w} \in \Sigma^{d\ell} \text{ such that } \delta(q, \mathbf{w}) = \bar{q} \text{ and } T(q, \mathbf{w}) = g \Rightarrow [\mathbf{w}]_k \equiv r \pmod{m}\}.$$

An equivalent definition for  $d_{g,\ell}^{q\bar{q}}$  is the greatest common divisor of all differences of numbers  $[\mathbf{w}]_k$  corresponding to paths of length  $d\ell$  from  $q$  to  $\bar{q}$  with weight  $g$ . We will use one of the two definitions depending on the situation. Our next goal is to show that  $d_{g,\ell}^{q\bar{q}}$  converges to some  $d(q, \bar{q})$  for all  $g \in G$ .

**Lemma 2.17.** *Let  $q, \bar{q} \in Q$ . Then there exists  $d(q, \bar{q}) := \lim_{\ell \rightarrow \infty} d_{id,\ell}^{q\bar{q}}$  and for all  $\ell \geq m_0, g \in \Delta$  we have  $d(q, \bar{q}) \mid d_{g,\ell}^{q\bar{q}}$  and there exists  $m'_0$  (not depending on  $q, \bar{q}$ ) such that for all  $\ell \geq m'_0, g \in \Delta$  we have  $d_{g,\ell}^{q\bar{q}} = d(q, \bar{q})$ .*

*Proof.* We start by showing that for all  $\ell, \ell' \geq m_0, q, \bar{q} \in Q$  and  $g_1, g_2 \in G$  it holds that  $d_{g_1, \ell + \ell'}^{q\bar{q}} \mid d_{g_2, \ell'}^{q\bar{q}}$ . Theorem 2.7 shows that there exists  $\mathbf{w} \in \Sigma^{d\ell'}$  such that  $\delta(q, \mathbf{w}) = q$  and  $T(q, \mathbf{w}) = g_1 \cdot g_2^{-1}$ . We find

$$\begin{aligned} d_{g_1, \ell + \ell'}^{q\bar{q}} &= \gcd(\{[\mathbf{w}_1]_k - [\mathbf{w}_2]_k : \mathbf{w}_i \in \Sigma^{d(\ell + \ell')} \text{ with } \delta(q, \mathbf{w}_i) = \bar{q}, T(q, \mathbf{w}_i) = g_1\}) \\ &\mid \gcd(\{[\mathbf{w}\mathbf{w}_1]_k - [\mathbf{w}\mathbf{w}_2]_k : \mathbf{w}_i \in \Sigma^{d\ell'} \text{ with } \delta(q, \mathbf{w}\mathbf{w}_i) = \bar{q}, T(q, \mathbf{w}\mathbf{w}_i) = g_1\}) \\ &= \gcd(\{[\mathbf{w}_1]_k - [\mathbf{w}_2]_k : \mathbf{w}_i \in \Sigma^{d\ell} \text{ with } \delta(q, \mathbf{w}_i) = \bar{q}, T(q, \mathbf{w}_i) = g_2\}) \\ &= d_{g_2, \ell}^{q\bar{q}} \end{aligned}$$

Let  $d_g^{q\bar{q}}$  denote the minimal value of  $d_{g,\ell}^{q\bar{q}}$  for all  $\ell$  (we choose  $\ell_0 \geq m_0$  such that  $d_{g,\ell_0}^{q\bar{q}} = d_g^{q\bar{q}}$ ) and since

$$d_g^{q\bar{q}} = d_{g, \ell + \ell_0}^{q\bar{q}} \leq d_{g, \ell_0}^{q\bar{q}} = d_g^{q\bar{q}}$$

there exists  $m'_0$  such that for all  $\ell \geq m'_0$  it holds that  $d_{g,\ell}^{q\bar{q}} = d_g^{q\bar{q}}$ . It also follows directly that  $d_{g_1}^{q\bar{q}} = d_{g_2}^{q\bar{q}}$  for all  $g_1, g_2 \in \Delta$  and the result follows directly.  $\square$

Let  $\mathbf{w} \in \Sigma^\ell$  and suppose  $\delta(q, \mathbf{w}) = \bar{q}, T(q, \mathbf{w}) = g$ , we then denote  $r_\ell^{q\bar{q}}(g) := [\mathbf{w}]_k \pmod{d(q, \bar{q})}$  as we see directly that this definition does not depend on the choice of  $\mathbf{w}$ , but only on  $q, \bar{q}$  and  $g$ . More generally, it follows that for  $\mathbf{w} \in \Sigma^{d\ell}, \delta(q, \mathbf{w}) = \bar{q}$  we have

$$r_\ell^{q\bar{q}}(T(q, \mathbf{w})) \equiv [\mathbf{w}]_k \pmod{d(q, \bar{q})}.$$

**Remark.** *By assumption any word containing a synchronizing word is again synchronizing, and by assumption  $\mathcal{T}_A$  is synchronizing and strongly connected, there exists a minimal  $\ell_0 = \ell_0(A)$  such that for all  $q \in Q$  there exists  $\mathbf{w}_q \in \Sigma^{\ell_0}$  with*

$$\forall \bar{q} \in Q : \delta(\bar{q}, \mathbf{w}_q) = q.$$

We find an important restriction on  $d(q, \bar{q})$ .

**Lemma 2.18.** *We have for every  $q, \bar{q} \in Q$*

$$k^{\ell_0} (k^d - 1) \equiv 0 \pmod{d(q, \bar{q})}.$$

*Proof.* We fix  $q, \bar{q}$  and start by considering the case where we concatenate a word  $\mathbf{w}$  with some words  $\mathbf{w}_1, \mathbf{w}_2$  from left and right, respectively, such that the weight does not change: Let  $\mathbf{w} \in \Sigma^{m_2}, \mathbf{w}_1, \mathbf{w}_2 \in \Sigma^{dm_1}$  ( $m_1 \geq m_0, m_2 \geq m'_0$ ) such that  $\delta(q, \mathbf{w}_1) = q, T(q, \mathbf{w}_1) = id$  and  $\delta(\bar{q}, \mathbf{w}_2) = \bar{q}, T(\bar{q}, \mathbf{w}_2) = id$ , see Theorem 2.7. We find by simple computations and Lemma 2.17

$$\begin{aligned} r_{m_2+dm_1}^{q\bar{q}}(T(q, \mathbf{w}_1\mathbf{w})) &= r_{dm_1+m_2}^{q\bar{q}}(T(q, \mathbf{w}\mathbf{w}_2)) \\ [\mathbf{w}_1]_k k^{m_2} + [\mathbf{w}]_k &\equiv [\mathbf{w}]_k k^{dm_1} + [\mathbf{w}_2]_k \pmod{d(q, \bar{q})} \\ [\mathbf{w}_1]_k k^{m_2} - [\mathbf{w}_2]_k &\equiv [\mathbf{w}]_k (k^{dm_1} - 1) \pmod{d(q, \bar{q})}. \end{aligned}$$

The left side of the last equation only depends on  $\mathbf{w}_1, \mathbf{w}_2$  and  $m_2$ . By choosing  $m_2 \geq \ell_0 + 1$ ,  $\mathbf{w} = 0\mathbf{w}_{\bar{q}}$  or  $\mathbf{w} = 1\mathbf{w}_{\bar{q}}$  and – if  $\mathbf{w}$  is not of sufficient length – adding zeros from the left we find

$$\begin{aligned} [\mathbf{w}_{\bar{q}}]_k (k^{dm_1} - 1) &\equiv (k^{\ell_0} + [\mathbf{w}_{\bar{q}}]_k) (k^{dm_1} - 1) \pmod{d(q, \bar{q})} \\ k^{\ell_0} (k^{dm_1} - 1) &\equiv 0 \pmod{d(q, \bar{q})}. \end{aligned}$$

Comparing the results for  $m_1$  and  $m_1 + 1$  completes the proof.  $\square$

This allows us to decompose  $d(q, \bar{q})$  into two co-prime factors  $d'(q, \bar{q}), d''(q, \bar{q})$  such that  $d'(q, \bar{q}) | (k^d - 1)$  and  $d''(q, \bar{q}) | k^{\ell_0}$ . On the one hand  $d''(q, \bar{q})$  corresponds to restrictions on how you can reach the state  $\bar{q}$ , on the other hand  $d'(q, \bar{q})$  corresponds to some restrictions mod  $k^d - 1$ , which is of greater interest for us.

**Lemma 2.19.** *For every  $q, \bar{q} \in Q$  it holds that  $d'(A) := d'(q_0, q_0) = d'(q, \bar{q})$ .*

*Proof.* Since we have already gained some knowledge about  $d(q, \bar{q})$  we can adapt Lemma 2.17 to our needs.

Let  $\ell, \ell_1, \ell_2 \geq m'_0, q_1, q_2, \bar{q}_1, \bar{q}_2 \in Q$  and paths  $\mathbf{w} \in \Sigma^{d\ell_1}$  from  $q_1$  to  $\bar{q}_1$  with weight  $id$ ,  $\mathbf{w}' \in \Sigma^{d\ell_2}$  from  $\bar{q}_2$  to  $q_2$  with weight  $id$ . It follows, as in Lemma 2.17,

$$\begin{aligned} d(q_1, q_2) &= d_{id, \ell_1 + \ell_2}^{q_1, q_2} = \gcd(\{[\mathbf{w}_1]_k - [\mathbf{w}_2]_k : \mathbf{w}_i \in \Sigma^{d(\ell_1 + \ell_2)} \text{ with } \delta(q_1, \mathbf{w}_i) = q_2, T(q_1, \mathbf{w}_i) = id\}) \\ &| \gcd(\{[\mathbf{w}\mathbf{w}_1\mathbf{w}']_k - [\mathbf{w}\mathbf{w}_2\mathbf{w}']_k : \\ &\quad \mathbf{w}_i \in \Sigma^{d\ell} \text{ with } \delta(q_1, \mathbf{w}\mathbf{w}_i\mathbf{w}') = q_2, T(q_1, \mathbf{w}\mathbf{w}_i\mathbf{w}') = id\}) \\ &= \gcd(\{k^{d\ell_2}([\mathbf{w}_1]_k - [\mathbf{w}_2]_k) : \mathbf{w}_i \in \Sigma^{d\ell} \text{ with } \delta(\bar{q}_1, \mathbf{w}_i) = \bar{q}_2, T(\bar{q}_1, \mathbf{w}_i) = id\}) \\ &= k^{d\ell_2} d_{id, \ell}^{\bar{q}_1, \bar{q}_2} = k^{d\ell_2} d(\bar{q}_1, \bar{q}_2) \end{aligned}$$

for arbitrary  $q_1, q_2, \bar{q}_1, \bar{q}_2$ . Thus we find  $d'(q_1, q_2) | d'(\bar{q}_1, \bar{q}_2)$ . As changing the order of some state  $q \in Q$  does not change  $d(\mathcal{T}_A)$  ( $id \mapsto \sigma_q \circ id \circ \sigma_q^{-1}$ ), we find that  $d'$  only depends on  $A$ .  $\square$

As mentioned before, we are more interested in  $d'$  and define  $s_\ell^{q\bar{q}}(g) := r_\ell^{q\bar{q}}(g) \pmod{d'}$ . We find the following important properties:

**Lemma 2.20.** *For all  $q_1, q_2, q_3 \in Q, m_1, m_2 \geq m'_0$ , and  $g_1, g_2 \in G$  follows*

$$s_{d(m_1+m_2)}^{q_1, q_3}(g_1 \cdot g_2) = s_{dm_1}^{q_1, q_2}(g_1) + s_{dm_2}^{q_2, q_3}(g_2).$$

*Proof.* We find by Theorem 2.7 that there exists  $\mathbf{w}_1 \in \Sigma^{dm_1}$  and  $\mathbf{w}_2 \in \Sigma^{dm_2}$  such that  $\delta(q_1, \mathbf{w}_1) = q_2, T(q_1, \mathbf{w}_1) = g_1$  and  $\delta(q_2, \mathbf{w}_2) = q_3, T(q_2, \mathbf{w}_2) = g_2$ . This gives,

$$\begin{aligned} s_{dm_1}^{q_1, q_2}(g_1) &\equiv [bfw_1]_k \pmod{d'} \\ s_{dm_2}^{q_2, q_3}(g_2) &\equiv [bfw_2]_k \pmod{d'} \\ s_{d(m_1+m_2)}^{q_1, q_3}(g_1 \cdot g_2) &\equiv [bfw_1]_k k^{dm_2} + [bfw_2]_k \pmod{d'}. \end{aligned}$$

Thus, the lemma follows directly from the fact that  $k^{dm_2} \equiv 1 \pmod{k^d - 1}$ .  $\square$

**Lemma 2.21.** *There exists  $k_0 \in \mathbb{N}$  such that for every  $\ell \geq m'_0$  we have  $s_{d\ell}^{q_0 q_0}(id) = 0 \Leftrightarrow \ell \equiv 0 \pmod{k_0}$ .*

*Proof.* We find by Lemma 2.20  $s_{d(\ell_1+\ell_2)}^{q_0 q_0}(id) = s_{d\ell_1}^{q_0 q_0}(id) + s_{d\ell_2}^{q_0 q_0}(id) \pmod{d'}$ . The statement follows by the same arguments as we used in the proof of Lemma 2.9.  $\square$

**Remark.** *One can actually prove that the Lemma above holds for all  $\ell \in \mathbb{N}$  for which  $s_{d\ell}^{q_0 q_0}(id)$  is properly defined. Furthermore,  $k_0 | d'$  holds.*

We further follow the ideas of Section 2.3 and find the following result.

**Lemma 2.22.** *There exists a naturally induced transducer  $\overline{\mathcal{T}}_A$  such that  $\overline{s}_{dk_0 m'_0}^{q_0 \overline{q}}(id) = 0$  holds for all  $\overline{q} \in \overline{Q}$ .*

*Proof.* We want to find some permutations  $\sigma_q$  such that applying them gives an induced transducer with the desired properties. Let  $q \neq q_0$ , take  $\mathbf{w} \in \Sigma^{dk_0 m'_0}$  such that  $[\mathbf{w}]_k \equiv 0 \pmod{d'}, \delta(q_0, \mathbf{w}) = q$  and define  $\sigma_q = T(q_0, \mathbf{w})$ . Without loss of generality we restrict ourself to  $\mathbf{w} = \mathbf{w}'\mathbf{w}_q$ , where  $\mathbf{w}_q$  is again a synchronizing word – we can easily choose a suitable  $\mathbf{w}'$ .

By choosing  $\sigma_{q_0} = id$  we find by using Proposition 2.6 that  $\overline{T}(\overline{q_0}, \mathbf{w}) = T(q_0, \mathbf{w}) \cdot T(q_0, \mathbf{w})^{-1} = id$ .

Note that applying these permutations does not change the property  $id \in G_{q_0 q}(0)$  used in Section 2.3  $\square$

We consider from now on only naturally induced transducers such that  $s_{k_0 m'_0}^{q_0 q}(id) = 0$  holds for all  $q \in Q$ .

We find a result similar to the one in Section 2.3 concerning  $M_{q\overline{q}, \ell}$ , which shows that  $s_{d\ell}^{q\overline{q}}(g)$  only depends on  $\ell \pmod{k_0}$  and  $g$ .

**Proposition 2.23.** *For all  $q, \overline{q} \in Q, g \in G$  and  $\ell \in \mathbb{N}, \ell \geq m'_0$  we have*

$$s_{\ell}^{q\overline{q}}(g) = s_{\ell \pmod{k_0}}(g),$$

where  $s_{\ell}(g) := s_{\ell \pmod{k_0}}(g) := s_{k_0 m'_0 + (\ell \pmod{k_0})}^{q_0 q_0}(g)$ .

*Proof.* We find by Lemma 2.20 that

$$0 = s_{2k_0 m'_0}^{q_0 q_0}(id) = s_{k_0 m'_0}^{q_0 q}(id) + s_{k_0 m'_0}^{q q_0}(id) = s_{k_0 m'_0}^{q q_0}(id).$$

Let us now assume  $\ell \geq 5k_0 m'_0$ . We have

$$s_{\ell}^{q\overline{q}}(g) = s_{k_0 m'_0}^{q q_0}(id) k^{\ell - k_0 m'_0} + s_{\ell - 2k_0 m'_0}^{q_0 q_0}(g) k^{k_0 m'_0} + s_{k_0 m'_0}^{q_0 \overline{q}}(id)$$

$$= s_{\ell-2k_0m_0}^{q_0q_0}(g) = s_{k_0m_0+(\ell \bmod k_0)}^{q_0q_0}(g) + s_{\ell-(\ell \bmod k_0)-3k_0m_0}^{q_0q_0}(id) = s_\ell(g).$$

Let now  $\ell \geq m'_0$ . We find

$$s_\ell^{q\bar{q}}(g) = s_\ell^{q\bar{q}}(g) + 5s_\ell^{q\bar{q}}(id) = s_{\ell+5k_0m'_0}^{q\bar{q}}(g) = s_{\ell \bmod k_0}(g),$$

which finishes the proof.  $\square$

We define  $G_\ell := \{g \in G : s_0(g) = \ell\}$ .

**Corollary 2.24.**  $G_0$  is a subgroup of  $G$ .

*Proof.* The statement follows directly by Lemma 2.20.  $\square$

Thus we can prove now Theorem 2.16.

*Proof of Theorem 2.16:* Of course we choose  $k_0, m'_0, \ell_0, d$  and  $G_0$  as we defined them throughout this section. To prove the first part we choose  $g'_0$  as an arbitrary element of  $G_1$  and the proof follows by the same arguments as in the proof of Theorem 2.7. The second part is just the result of Lemma 2.19.  $\square$

**2.5. Reduction to special naturally induced transducers.** We want to show how to reduce the general setting of an automatic sequence to automata with special properties.

**Proposition 2.25.** *Let  $\mathbf{a} = (a_n)_{n \in \mathbb{N}}$  be a  $k$ -automatic sequence. There exists  $p \in \mathbb{N}$  and  $A = (Q', \{0, \dots, k^p - 1\}, \delta', q'_0)$  such that  $\mathbf{a}$  is generated by  $A$  and for every automaton  $A_i$  that is a restriction to a final component of  $A$ , it holds that  $d(A_i) = k_0(A_i) = 1$ .*

*Proof.* Let  $\bar{A} = (\bar{Q}, \{0, \dots, k - 1\}, \bar{\delta}, \bar{q}_0)$  be an automaton that generates  $\mathbf{a}$ . We can assume without loss of generality that  $\bar{\delta}(\bar{q}_0, 0) = \bar{q}_0$  and every state  $\bar{q} \in \bar{Q}$  is reachable from  $\bar{q}_0$ . Let  $\bar{Q}_1, \dots, \bar{Q}_\ell$  be the final components of  $\bar{A}$ . We let  $\bar{A}_i$  denote the automaton that corresponds to the restriction of  $\bar{A}$  to  $\bar{Q}_i$ , with arbitrary initial state.

We define  $p = \text{lcm}(d(\bar{A}_1)k_0(\bar{A}_1), \dots, d(\bar{A}_\ell)k_0(\bar{A}_\ell))$  and note that  $d(\bar{A}_i)$  and  $k_0(\bar{A}_i)$  do not depend on the initial state of  $\bar{A}_i$  – compare Theorem 2.7 and Theorem 2.16. The idea is now to take the “ $p$ -th power” of  $\bar{A}$  which will give the desired property.

Therefore we define  $A = (\bar{Q}, \{0, \dots, k^p - 1\}, \delta', \bar{q}_0)$  where  $\delta'$  is the extension of  $\bar{\delta}$  to  $\{0, \dots, k - 1\}^p \cong \{0, \dots, k^p - 1\}$  from letters to words of length  $p$ . We see that  $A$  generates  $\mathbf{a}$  as  $\bar{\delta}(\bar{q}_0, 0) = \bar{q}_0$  ensures that adding leading zeros does not change the output (of  $\bar{A}$ ).

One easily observes that every final component of  $A$  is contained in a final component of  $\bar{A}$ : The strongly connected components of a directed graph form a new directed acyclic graph where every node corresponds to a strongly connected component. There exists an edge between two such nodes if and only if there is an edge between two states of these strongly connected components. Every final component of  $A$  is part of a final component of  $\bar{A}$ , since there exists a path (whose length is divisible by  $p$ ) from every strongly connected component, which is not closed under  $\bar{\delta}$  to a final component. Therefore, we can restrict ourself to consider only a final component  $Q'_i$ .

It remains to show  $d(A_i) = k_0(A_i) = 1$ . For this purpose, we want to describe how a naturally induced transducer of  $A_i$  looks like. Let  $\mathcal{T}_{\bar{A}_i} = (Q, \{0, \dots, k - 1\}, \delta, q_0, \Delta, \lambda)$  be a naturally

induced transducer of  $\bar{A}_j$ . As  $d(\bar{A}_j)k_0(\bar{A}_j)|p$  we know by Theorem 2.7 that for all  $q_1, q_2 \in Q$  and  $g \in G$  there exists  $\mathbf{w} \in \{0, \dots, k^p - 1\}^*$  such that  $\delta(q_1, \mathbf{w}) = q_2$  and  $T(q_1, \mathbf{w}) = g$ . This means, in particular, that there exists  $\mathbf{w} \in \{0, \dots, k^p - 1\}^*$  such that  $\delta(q_1, \mathbf{w}) = q_2$  and  $T(q_1, \mathbf{w}) = id$ . Thus Proposition 2.5 shows that we can describe the states of  $A_i$  as the elements corresponding to some coordinates of  $\mathcal{T}_{\bar{A}_j}$ :  $\pi_i(q_1)$  and  $\pi_i(q_2)$  belong to the same strongly connected component for all  $q_1, q_2 \in Q$  as  $\mathbf{w}$  satisfies  $\delta'(\pi_i(q_1), \mathbf{w}) = \pi_i(q_2)$ . This implies that

$$Q'_i = \{\pi_\ell(q) | q \in Q, \ell \in I\},$$

holds for some index set  $I \subseteq \{1, \dots, n_0\}$ . We claim that  $\mathcal{T} := (Q_I, \Sigma^p, \tilde{\delta}_I, (q_0)_I, G_I, \tilde{\lambda}_I)$  provides a naturally induced transducer of  $A_i$ . Here we again denote by  $\tilde{\delta}$  the extension of  $\delta$  from letters to words,  $\tilde{\lambda}$  coincides with  $T$  for words of length  $p$  (for  $\mathcal{T}_A$ ) and  $G$  is defined as in Theorem 2.7. Furthermore, we use  $(\cdot)_I$  to denote the projection to the coordinates of  $I$ . Theorem 2.7 assures that  $\mathcal{T}_A$  is an induced transducer. To see that  $\mathcal{T}_A$  is synchronizing we just have to take a synchronizing word  $\mathbf{w}_{q_0}$  of  $\mathcal{T}_A$  and add leading zeros such that it is a word whose length is divisible by  $p$ .

However Property 6) may not hold, but we already discussed in Section 2.1 that this can be fixed easily.

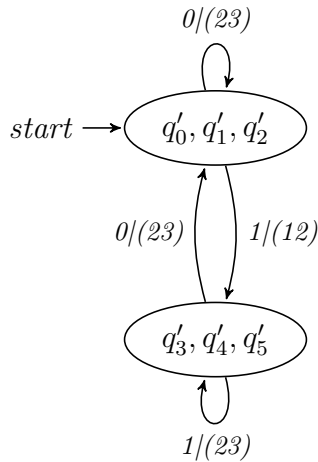
By the construction of  $\mathcal{T}$  we see easily that  $d(A_i) = k_0(A_i) = 1$  holds.  $\square$

This Proposition can be simplified under suitable conditions.

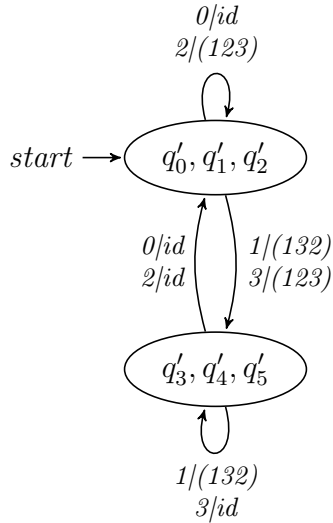
**Corollary 2.26.** *Let  $\bar{A} = (Q', \Sigma, \delta', q'_0)$  be a strongly connected automaton, such that  $\delta'(q'_0, 0) = q'_0$ . Then there exists a strongly connected automaton  $A$  such that  $d(A) = k_0(A) = 1$  and the automatic sequences generated by  $A$  and  $\bar{A}$  coincide.*

*Proof.* We consider the proof of Proposition 2.25 and see directly that  $A$  – as considered in the proof of Proposition 2.25 – has only one strongly connected component. It just remains to note that the automatic sequences generated by  $A$  and  $\bar{A}$  coincide since adding leading zeros does not change the output.  $\square$

**Example.** *We continue the example from Section 2.3. The naturally induced transducer we considered was*



and we have already seen that  $d = 2, k_0 = 1$ . Thus we consider the automaton that is the "square" of the original automaton, i.e., we concatenate two steps of the original automaton to one step of the new automaton, and the corresponding naturally induced transducer:



We find then easily that for this transducer  $d = k_0 = 1$  holds.

### 3. REDUCTION OF THEOREM 1.2 AND THEOREM 1.3

We reduce in this section the two main Theorems 1.2 and 1.3 to statements concerning naturally induced transducers.

**3.1. Möbius Randomness Principle.** First we present a lemma that reduces the Sarnak conjecture for the symbolic dynamical system associated with an automatic sequence to a Möbius randomness law for (possibly different) automatic sequences.

**Lemma 3.1.** *Suppose that for every automatic sequence  $(a_n)_{n \in \mathbb{N}}$  with values in  $\mathbb{C}$*

$$(2) \quad \sum_{n \leq N} \mu(n) a_{n+r} = o(N),$$

*uniformly for  $r \in \mathbb{N}$ . Then Theorem 1.2 holds.*

*Proof.* We now fix one automatic sequence  $\mathbf{b} = (b_n)_{n \in \mathbb{N}}$  that takes values in  $\mathbb{C}$ . Let  $(X, S)$  be the symbolic dynamical system associated with  $\mathbf{b}$ . [7, Lemma 2.1] shows that it is sufficient to show that for every  $j \geq 1$  and every  $g : \mathbb{C}^j \rightarrow \mathbb{C}$  it holds that

$$\sum_{n \leq N} \mu(n) g(b_{n+r}, b_{n+r+1}, \dots, b_{n+r+j-1}) = o(N)$$

uniformly for  $r \in \mathbb{N}$ . However, [1, Theorem 5.4.4, Lemma 5.2.6 and Lemma 4.3.9] shows that  $(b_{n+\ell})_{n \in \mathbb{N}}$  is again an automatic sequence for all  $\ell \in \mathbb{N}$ , the product of two automatic sequences is again an automatic sequence, and therefore the product of  $j$  automatic sequences is again an automatic sequence, and also that  $(g(b_n))_{n \in \mathbb{N}}$  is again an automatic sequence. Therefore,  $a_n = g(b_{n+r}, b_{n+r+1}, \dots, b_{n+r+j-1})$  defines again an automatic sequence and hence the statement follows.  $\square$



To further reduce this result we need some ideas of representation theory. We cover the most important definitions and notations briefly. A  $d$  dimensional representation  $D$  is a continuous homomorphism  $D : G \rightarrow U_d$ , where we let  $U_d$  denote the group of unitary  $d \times d$  matrices over  $\mathbb{C}$ . A representation  $D$  is called irreducible if there exists no non-trivial subspace  $V \subset U_d$  such that  $D(g)V \subseteq V$  holds for all  $g \in G$ . Furthermore, we will use a scalar-product for functions  $f_1, f_2 : G \rightarrow \mathbb{C}$

$$\langle f_1, f_2 \rangle := \int f_1 \overline{f_2} d\boldsymbol{\mu} = \frac{1}{|G|} \sum_{g \in G} f_1(g) \overline{f_2(g)},$$

where  $\boldsymbol{\mu}$  denotes the Haar measure. In Section 4, we show the following result.

**Proposition 3.2.** *Let  $\mathcal{T}_A$  be a naturally induced transducer by  $A$  and suppose that  $d(A) = 1, k_0(A) = 1$  holds. Let, furthermore,  $D$  be a unitary irreducible representation of  $G$ ,  $r, \lambda_1, \lambda_2 \in \mathbb{N}, 0 < b < k^{\lambda_1}$  and  $m < k^{\lambda_2}$ . It holds*

$$(3) \quad \left\| \sum_{\substack{n < N \\ n \equiv m \pmod{k^{\lambda_2}}} \chi_{1/k^{\lambda_1}} \left( \frac{n+r-bk^{\nu-\lambda_1}}{k^\nu} \right) D(T(q_0, (n+r)_k)) \mu(n) \right\|_F = o(N),$$

uniformly for  $r$  where  $\nu$  is the unique integer satisfying  $k^{\nu-1} \leq N < k^\nu$ .

**Remark.**  $n \equiv m \pmod{k^{\lambda_2}}$  obviously fixes the last  $\lambda_2$  digits of  $n$ . Furthermore,  $\chi_{1/k^{\lambda_1}}(\cdot)$  fixes the digits  $\nu - \lambda_1, \dots, \nu - 1$  of  $n$ .

Our goal of this subsection is to show that Proposition 3.2 implies Theorem 1.2.

We fix any  $\varepsilon > 0$  and need to show that there exists  $N_0$  such that for  $N \geq N_0(\varepsilon)$

$$|S(r, N)| \leq \varepsilon N$$

holds for all  $r$ , where

$$S(r, N) := \sum_{n < N} \mu(n) a_{n+r}.$$

Let us first present the most important ideas for this proof. Obviously, we can only work with naturally induced transducers, when we restrict ourself to the final components of  $A$ . Therefore we fix some of the first digits of  $(n+r)$  which allows us to work with a strongly connected component of  $Q'$ . Thereafter, we also fix the last digits of  $(n+r)$  to fix  $\delta(q, (n+r)_k)$ . It then remains to find suitable estimates for  $T(q, \cdot)$ .

**Remark.** *It would be very convenient to apply [7, Lemma 3.2] to only work with strongly connected automata, but this Lemma does not provide uniformity in  $r$ .*

More precisely, Proposition 2.25 shows that there exists a DFAO  $A = (Q', \Sigma, \delta', q'_0, \tau)$  such that  $\Sigma = \{0, \dots, k-1\}$ ,  $\delta'(q'_0, 0) = q'_0$  and  $a_n = \tau(\delta'(q'_0, (n)_k))$ . Let  $Q'_1, \dots, Q'_\ell$  be the strongly connected components of  $Q'$  that are closed under  $\delta'$ . Furthermore, let  $A_i$  be the restriction of  $A$  to  $Q'_i$ . Proposition 2.25 states that  $d(A_i) = k_0(A_i) = 1$  holds.

We define the naturally induced transducers for these components by

$$\mathcal{T}_{A_i} = (Q_i, \Sigma, \delta_i, \Delta_i, \lambda_i)$$

and let  $\mathbf{w}_i$  be a synchronizing word of  $\mathcal{T}_{A_i}$  (we intentionally avoid to define the initial state here).

**Lemma 3.3.** *The set*

$M := \{n \in \mathbb{N} \mid \exists \mathbf{v}_1, \mathbf{v}_2 \in \Sigma^* : (n)_k = \mathbf{v}_1 \mathbf{v}_2 \wedge \forall q' \in Q' : \delta'(q', \mathbf{v}_1) \in \cup Q'_i \wedge \forall i : \mathbf{w}_i \text{ is a subword of } \mathbf{v}_2\}$   
*has density 1.*

The motivation for introducing the set  $M$  is that whenever you read a word corresponding to some  $n \in M$ , you end up in one final component  $Q'_i$  – this is achieved by  $\mathbf{v}_1$ . The purpose of  $\mathbf{v}_2$  is not so obvious but will be more apparent shortly.

**Remark.** *Let  $n \in M$ . By the definition of  $M$  it follows that the considered properties hold also for unreduced digital representations of  $n$  (i.e., with leading zeros).*

As the proof of Lemma 3.3 is rather technical, we postpone it to Section 5.

We assume for simplicity that  $|a_n| \leq 1$  holds for all  $n \in \mathbb{N}$ . We let  $\nu$  denote the unique integer satisfying  $k^{\nu-1} \leq N < k^\nu$ .

The first idea is to remove the digits of  $n+r$  at positions  $\nu, \nu+1, \dots$ , i.e.,  $(\lfloor (n+r)/k^\nu \rfloor)_k$ , and restrict ourself to  $(n+r) \bmod k^\nu$  as these integers form (at most) two intervals with combined length at least  $k^{\nu-1}$ . Therefore, we want to detect the digits of  $n+r$  at the positions  $\nu-\lambda_1, \dots, \nu-1$  for some fixed value  $\lambda_1$  – only depending on  $\varepsilon$  – that we will choose shortly:

$$\begin{aligned} \left| \sum_{n < N} \mu(n) a_{n+r} \right| &= \left| \sum_{n < N} \sum_{b < k^{\lambda_1}} \chi_{k^{-\lambda_1}} \left( \frac{n+r - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) a_{n+r} \right| \\ &\leq \sum_{b < k^{\lambda_1}} \left| \sum_{n < N} \chi_{k^{-\lambda_1}} \left( \frac{n+r - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) a_{n+r} \right| \\ &\leq \sum_{\substack{b < k^{\lambda_1} \\ b \in M}} \left| \sum_{n < N} \chi_{k^{-\lambda_1}} \left( \frac{n+r - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) a_{n+r} \right| + \sum_{\substack{b < k^{\lambda_1} \\ b \notin M}} k^{\nu-\lambda_1}. \end{aligned}$$

Choosing  $\lambda_1 (< \nu)$  such that  $|\{b < k^{\lambda_1} \mid b \notin M\}| \leq \frac{\varepsilon}{3k} k^{\lambda_1}$  gives

$$|S(r, N)| \leq \sum_{\substack{b < k^{\lambda_1} \\ b \in M}} |S_1(b, r, N)| + \frac{\varepsilon}{3} N$$

where

$$S_1 := S_1(b, r, N) := \sum_{n < N} \chi_{k^{-\lambda_1}} \left( \frac{n+r - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) a_{n+r}.$$

We rewrite  $r = r_1 + k^\nu r_2$  with  $r_1 < k^\nu$  and  $r_1, r_2 \in \mathbb{N}$ , and split the sum over  $n$  into two parts. We let  $(n)_k^t$  denote the unique word  $\mathbf{w}$  of length  $t$  such that  $[\mathbf{w}]_k \equiv n \pmod{k^t}$  and find

$$|S_1| \leq \left| \sum_{\substack{n < N \\ n+r_1 < k^\nu}} \chi_{k^{-\lambda_1}} \left( \frac{n+r - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) \tau(\delta'(q'_0, (n+r)_k)) \right|$$

$$\begin{aligned}
& + \left| \sum_{k^\nu - r_1 \leq n < N} \chi_{k^{-\lambda_1}} \left( \frac{n + r - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) \tau(\delta'(q'_0, (n+r)_k)) \right| \\
& = \left| \sum_{\substack{n < N \\ n < k^\nu - r_1}} \chi_{k^{-\lambda_1}} \left( \frac{n + r - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) \tau(\delta'(q'_0, (r_2)_k (b)_k^{\lambda_1} (n+r)_k^{\nu-\lambda_1})) \right| \\
& + \left| \sum_{k^\nu - r_1 \leq n < N} \chi_{k^{-\lambda_1}} \left( \frac{n + r - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) \tau(\delta'(q'_0, (r_2 + 1)_k (b)_k^{\lambda_1} (n+r)_k^{\nu-\lambda_1})) \right|.
\end{aligned}$$

Thus we find

$$(4) \quad |S_1| \leq |S_2(b, r_1, \min(N, k^\nu - r_1), \delta'(q'_0, (r_2)_k))| + |S_2(b, r_1, N, \delta'(q'_0, (r_2 + 1)_k))| \\ + |S_2(b, r_1, k^\nu - r_1, \delta'(q'_0, (r_2 + 1)_k))|,$$

where

$$S_2(b, r', N', q') := \sum_{n < N'} \chi_{k^{-\lambda_1}} \left( \frac{n + r' - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) \tau(\delta'(q', (b)_k^{\lambda_1} (n+r')_k^{\nu-\lambda_1}))$$

for  $N' + r' < k^\nu$ .

We simplify (4) to

$$(5) \quad |S_1(b, r, N)| \leq 3 \max_{\substack{r', N' \in \mathbb{N} \\ r' + N' < k^\nu}} \max_{q' \in Q'} |S_2(b, r', N', q')|$$

Now we work on an estimate for  $S_2$ . We can rewrite  $(b)_k^{\lambda_1} = \mathbf{v}_1 \mathbf{v}_2$  where  $\delta'(q', \mathbf{v}_1) \in Q'_i$  for some  $i$ .

We denote by  $q'_1 = \delta'(q', \mathbf{v}_1)$  and  $A(q'_1) = (Q'_i, \Sigma, \delta'_i, q'_1, \tau_i)$  the restriction of  $A$  to  $Q'_i$ , i.e.,  $\delta'_i = \delta'_i|_{Q'_i \times \Sigma}$ ,  $\tau_i = \tau|_{Q'_i}$ . Let  $\mathcal{T}_{A(q'_1)} = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$  be a naturally induced transducer of  $A(q'_1)$ . We define  $M_{\lambda_2}$  as the set of integers  $m$  such that  $(m)_{k^{\lambda_2}}$  is a synchronizing words of  $\mathcal{T}_{A(q'_1)}$  and note that for  $[\mathbf{w}_0]_k \in M_{\lambda_2}$ ,  $\mathbf{w} \in \Sigma^*$  it holds that  $\delta(q_0, \mathbf{w}\mathbf{w}_0) = \delta(q_0, \mathbf{w}_0)$ . [7, Lemma 2.2] shows that there exist  $c, \eta > 0$  such that  $k^{\lambda_2} - |M_{\lambda_2}| \leq ck^{(1-\eta)\lambda_2}$  holds for all  $\lambda_2 \in \mathbb{N}$ . Proposition 2.5 shows that

$$\begin{aligned}
|S_2(b, r', N', q')| & = \left| \sum_{n < N'} \chi_{k^{-\lambda_1}} \left( \frac{n + r' - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) \right. \\
& \quad \left. \tau(\pi_1(T(q_0, \mathbf{v}_2(n+r')_k^{\nu-\lambda_1})\delta(q_0, \mathbf{v}_2(n+r')_k^{\nu-\lambda_1}))) \right| \\
& = \left| \sum_{m < k^{\lambda_2}} \sum_{\substack{n < N' \\ n+r' \equiv m \pmod{k^{\lambda_2}}} \chi_{k^{-\lambda_1}} \left( \frac{n + r' - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) \right. \\
& \quad \left. \tau(\pi_1(T(q_0, \mathbf{v}_2(n+r')_k^{\nu-\lambda_1})\delta(q_0, \mathbf{v}_2(n+r')_k^{\nu-\lambda_1}))) \right|
\end{aligned}$$

$$\leq \sum_{m \in M_{\lambda_2}} \left| \sum_{\substack{n < N' \\ n \equiv m - r' \pmod{k^{\lambda_2}}} \chi_{k^{-\lambda_1}} \left( \frac{n + r' - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) \right. \\ \left. \tau(\pi_1(T(q_0, \mathbf{v}_2(n+r')_k^{\nu-\lambda_1})\delta(q_0, (m)_k^{\lambda_2}))) \right| + ck^{(1-\eta)\lambda_2} k^{\nu-\lambda_1-\lambda_2}.$$

We choose  $\lambda_2$  such that  $ck^{-\eta\lambda_2} \leq \frac{\varepsilon}{9k}$  (and we also assume that  $\lambda_1 + \lambda_2 < \nu$ ) and find

$$|S_2(b, r', N', q')| \leq \sum_{m \leq k^{\lambda_2}} \sum_{q'_1 \in \cup Q'_i} |S_3(b, r', N', m, \mathcal{T}_{A(q'_1)})| + \frac{\varepsilon N}{9} k^{-\lambda_1},$$

where

$$S_3(b, r', N', m, \mathcal{T}_{A(q'_1)}) := \sum_{\substack{n < N' \\ n \equiv m \pmod{k^{\lambda_2}}} \chi_{k^{-\lambda_1}} \left( \frac{n + r' - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) \\ \tau(\pi_1(T(q_0, \mathbf{v}_2(n+r')_k^{\nu-\lambda_1})\delta(q_0, (m)_k^{\lambda_2}))).$$

Note that  $\lambda_1$  and  $\lambda_2$  are independent of  $N$  and  $N'$ .

We define  $q_1 := \delta(q_0, (b)_k)$  and find  $T(q_0, (b)_k(n+r')_k^{\nu-\lambda_1}) = T(q_0, (b)_k) \circ T(q_1, (n+r')_k^{\nu-\lambda_1})$  as well as  $T(q_0, \mathbf{v}_2(n+r')_k^{\nu-\lambda_1}) = T(q_0, \mathbf{v}_2) \circ T(q_1, (n+r')_k^{\nu-\lambda_1})$ , since  $\mathbf{v}_2$  is synchronizing. This gives in total

$$T(q_0, \mathbf{v}_2(n+r')_k^{\nu-\lambda_1}) = T(q_0, \mathbf{v}_2) \circ T(q_0, (b)_k)^{-1} \circ T(q_1, (b)_k(n+r')_k^{\nu-\lambda_1}).$$

We define for  $\mathbf{w} \in \Sigma^*$  a function  $f_{\mathbf{w}, b} : G \rightarrow \mathbb{C}$

$$f_{\mathbf{w}, b}(\sigma) := \tau(\pi_1(T(q_0, \mathbf{v}_2) \circ (T(q_0, (b)_k)^{-1} \circ \sigma) \cdot (\delta(q_0, \mathbf{w}))))).$$

Thus we find  $\tau(\pi_1(T(q_0, \mathbf{v}_2(n+r')_k^{\nu-\lambda_1}) \cdot \delta(q_0, (m)_k^{\lambda_2}))) = f_{(m)_k^{\lambda_2}, b}(T(q_0, (b)_k(n+r')_k^{\nu-\lambda_1}))$ . We denote by  $\mathcal{F}(\mathcal{T}_{A(q'_1)}) := \{f_{\mathbf{w}, b} : \mathbf{w} \in \Sigma^*, b \in \mathbb{N}\}$ . Note that  $|\mathcal{F}(\mathcal{T}_{A(q'_1)})| \leq |Q| \cdot |G|$  holds, as  $\delta(q_0, \mathbf{w})$  can take at most  $|Q|$  and  $T(q_0, \mathbf{v}_2) \circ T(q_0, (b)_k)^{-1}$  can take at most  $|G|$  values. We find

$$S_3(b, r', N', m, \mathcal{T}_{A(q'_1)}) = \sum_{\substack{n < N' \\ n \equiv m \pmod{k^{\lambda_2}}} \chi_{k^{-\lambda_1}} \left( \frac{n + r' - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) f_{(m)_k^{\lambda_2}, b}(T(q_0, (n+r')_k)).$$

With  $T(n) := T(q_0, (n)_k)$  we find

$$|S_3(b, r', N', m, \mathcal{T}_{A(q'_1)})| \leq \max_{f \in \mathcal{F}(\mathcal{T}_{A(q'_1)})} \left| \sum_{\substack{n < N' \\ n \equiv m \pmod{k^{\lambda_2}}} \chi_{k^{-\lambda_1}} \left( \frac{n + r' - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) f(T(n+r')) \right|.$$

Consider a finite group  $G$ . It is well-known that there only exist finitely many equivalence classes of unitary and irreducible representations of  $G$  (see for example [35, Part I, Section 2.5]). The Peter-Weyl Theorem (see for example [27, Chapter 4, Theorem 1.2]) states that

the entry functions of irreducible representations (suitably renormalized) form an orthonormal basis of  $L^2(G)$ . Thus we can express any function  $f : G \rightarrow \mathbb{C}$  by these  $M_0$  entry functions:

Let  $D^{(m')} = (d_{ij}^{(m')})_{ij}$  be representations of the equivalence classes mentioned above and  $f : G \rightarrow \mathbb{C}$ . Then, there exist  $c_\ell$  such that

$$f(g) = \sum_{\ell < M_0} c_\ell d_{i_\ell j_\ell}^{(m_\ell)}(g)$$

holds for all  $g \in G$ . Thus we find

$$|S_3| \leq \max_{D \in \mathcal{D}_{q'_1}} c \left\| \sum_{\substack{n < N' \\ n \equiv m \pmod{k^{\lambda_2}}} \chi_{k^{-\lambda_1}} \left( \frac{n + r' - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) D(T(n + r')) \right\|_F,$$

where  $\|\cdot\|_F$  denotes the Frobenius norm,  $c$  is the maximum of all possible values of  $c_\ell \cdot M_0$  and  $\mathcal{D}_{q'_1}$  is the finite set of all possible appearing representations for  $\mathcal{F}(\mathcal{T}_{A(q'_1)})$ .

If we denote by  $T_{q'_1}$  the function  $T$  for  $\mathcal{T}_{A(q'_1)}$  we find in total

$$\left| \sum_{n < N} \mu(n) a_{n+r} \right| \leq \frac{2}{3} \varepsilon N + c \sum_{b < k^{\lambda_1}} \sup_{\substack{r', N' \in \mathbb{N} \\ N' + r' < k^\nu}} \sum_{m < k^{\lambda_2}} \sum_{q'_1 \in \cup Q'_i} \sum_{D \in \mathcal{D}_{q'_1}} \left\| \sum_{\substack{n < N' \\ n \equiv m \pmod{k^{\lambda_2}}} \chi_{k^{-\lambda_1}} \left( \frac{n + r' - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) D(T_{q'_1}(n + r')) \right\|_F.$$

Note that  $\mathcal{D}_{q'_1}$  is finite and only depends on the automatic sequence  $\mathbf{a}$ .

$$|S(r, N)| \leq \frac{2}{3} \varepsilon N + c \sum_{b < k^{\lambda_1}} \sum_{m < k^{\lambda_2}} \sum_{q'_1 \in \cup Q'_i} \sum_{D \in \mathcal{D}_{q'_1}} \sup_{r' \in \mathbb{N}} \max_{N' < k^\nu} \left\| \sum_{\substack{n < N' \\ n \equiv m \pmod{k^{\lambda_2}}} \chi_{k^{-\lambda_1}} \left( \frac{n + r' - bk^{\nu-\lambda_1}}{k^\nu} \right) \mu(n) D(T_{q'_1}(n + r')) \right\|_F,$$

where  $\lambda_1, \lambda_2$  and  $\mathcal{D}_{q'_1}$  only depend on  $\varepsilon$ . This finally proves the following proposition, which recapitulates the results of this subsection.

**Proposition 3.4.** *Assume Proposition 3.2 holds. Then Theorem 1.2 holds.*

**3.2. Prime Number Theorem.** We use the same notations as in the previous section and some similar ideas. We have seen that representations play an important role. For the prime number theorem we have to distinguish some special representations.

Recall that  $G_\ell := \{g \in G : s_0(g) = \ell\}$  for  $\ell = 0, \dots, d'(A) - 1$ .

**Lemma 3.5.** *There exist  $d'$  representations  $D_0, \dots, D_{d'-1}$  defined by*

$$D_\ell(g) := e \left( \frac{\ell \cdot s_0(g)}{d'} \right),$$

for  $\ell = 0, \dots, d' - 1$ , where  $s_0$  is defined by  $s_0(T(q, \mathbf{w})) \equiv [\mathbf{w}]_k \pmod{d'}$ .

*Proof.* The proof follows directly by Theorem 2.16 and Theorem 2.7.  $\square$

**Remark.** *These representations play an important role in this section.*

Note that  $D_\ell(T(q_0, \mathbf{w})) = e((\ell \cdot [\mathbf{w}]_k)/d')$  for  $\ell = 0, \dots, d' - 1, \mathbf{w} \in \Sigma^*$ .

In Section 4 we will prove the following Proposition 3.6 and in Section 5.2 deduce from it the more technical Lemma 3.7 below. The rest of this subsection is devoted to prove that this lemma implies Theorem 1.3.

**Proposition 3.6.** *Let  $D$  be a unitary, irreducible representation of  $G$  different from  $D_\ell$ . Then there exists some  $\eta > 0$  such that*

$$\left\| \frac{1}{k^\nu} \sum_{p < k^\nu} D(T(q_0, (p)_k)) e(-pt) \right\|_2 \ll k^{-\nu\eta}$$

*holds uniformly in  $t \in \mathbb{R}$ .*

**Lemma 3.7.** *Let  $\lambda \in \mathbb{N}$ ,  $\mathcal{T}$  be a naturally induced transducer with function  $T$ . For every  $g \in G$  there exists  $f_g$  such that*

$$\frac{1}{\pi(x; a, k^\lambda)} \sum_{\substack{p < x \\ p \equiv a \pmod{k^\lambda}}} \mathbf{1}_{[T(q_0, (p)_k) = g]} = f_g + o(1) \text{ for } x \rightarrow \infty$$

*holds (uniformly) for all  $a < k^\lambda$  such that  $(a, k) = 1$ .*

As mentioned before, the rest of this subsection is devoted to prove

**Proposition 3.8.** *Suppose Proposition 3.6 holds. Then Theorem 1.3 holds.*

*Proof.* We start as in the previous subsection and find by using Corollary 2.26 and Proposition 2.5 that

$$\begin{aligned} \frac{1}{\pi(x)} \sum_{p \leq x} \mathbf{1}_{[a_p = b]} &= \frac{1}{\pi(x)} \left( \sum_{\substack{a \leq k^\lambda \\ (a, k^\lambda) = 1}} \sum_{\substack{p \leq x \\ p \equiv a \pmod{k^\lambda}}} \mathbf{1}_{[a_p = b]} + O(k) \right) \\ &= \frac{1}{\pi(x)} \left( \sum_{\substack{a \leq k^\lambda \\ (a, k^\lambda) = 1}} \sum_{\substack{p \leq x \\ p \equiv a \pmod{k^\lambda}}} \mathbf{1}_{[\tau(\pi_1(T(q_0, (p)_k) \cdot \delta(q_0, (p)_k))) = b]} + O(k) \right), \end{aligned}$$

where the error term arises from primes that are not co-prime to  $k^\lambda$  and therefore not co-prime to  $k$ .

We note that  $(a, k^\lambda) = 1$  holds if and only if  $(\varepsilon_0(a), k) = 1$  where  $\varepsilon_0(x)$  denotes the least significant digit of  $x$  in base  $k$ . We denote again by  $M_\lambda := \{n < k^\lambda : (n, k) = 1 \text{ and } (n)_k \text{ is synchronizing}\}$ . One finds easily that  $|\{n \leq k^\lambda : (n, k) = 1, n \notin M_\lambda\}| = O(k^{\lambda(1-\eta)})$  for some  $\eta > 0$ , as in the previous section.

We fix  $\lambda$  for now to find the necessary estimates. Thereafter we let  $\lambda$  grow. This then gives the desired result.

$$\begin{aligned}
\frac{1}{\pi(x)} \sum_{p \leq x} \mathbf{1}_{[a_p=b]} &= \frac{1}{\pi(x)} \sum_{a \in M_\lambda} \sum_{\substack{p \leq x \\ p \equiv a \pmod{k^\lambda}}} \mathbf{1}_{[\tau(\pi_1(T(q_0, (p)_k) \cdot \delta(q_0, (p)_k)))=b]} \\
&\quad + \frac{\pi(x; 1, k^\lambda)}{\pi(x)} O(k^{\lambda(1-\eta)}) + O\left(\frac{k^\lambda}{\pi(x)}\right) \\
&= \frac{1}{\pi(x)} \sum_{a \in M_\lambda} \sum_{\substack{p \leq x \\ p \equiv a \pmod{k^\lambda}}} \mathbf{1}_{[\tau(\pi_1(T(q_0, (p)_k) \cdot \delta(q_0, (a)_k)))=b]} + O(k^{-\lambda\eta}) + O\left(\frac{k^\lambda}{\pi(x)}\right) \\
&= \frac{1}{\pi(x)} \sum_{q \in Q} \sum_{\substack{a \in M_\lambda \\ \delta(q_0, (a)_k) = q}} \sum_{\substack{p \leq x \\ p \equiv a \pmod{k^\lambda}}} \mathbf{1}_{[\tau(\pi_1(T(q_0, (p)_k) \cdot q))=b]} + O(k^{-\lambda\eta}) + O\left(\frac{k^\lambda}{\pi(x)}\right) \\
&= \frac{1}{\pi(x)} \sum_{q \in Q} \sum_{\substack{a < k^\lambda \\ (a, k) = 1}} \mathbf{1}_{[\delta(q_0, (a)_k) = q]} \sum_{\substack{p \leq x \\ p \equiv a \pmod{k^\lambda}}} \mathbf{1}_{[\tau(\pi_1(T(q_0, (p)_k) \cdot q))=b]} \\
&\quad + O(k^{-\lambda\eta}) + O\left(\frac{k^\lambda}{\pi(x)}\right).
\end{aligned}$$

Thus we are interested in the distribution of  $T(q_0, (p)_k)$  along primes in arithmetic progressions, which is covered by Lemma 3.7.

By writing

$$\sum_{\substack{p \leq x \\ p \equiv a \pmod{k^\lambda}}} \mathbf{1}_{[\tau(\pi_1(T(q_0, (p)_k) \cdot q))=b]} = \sum_{g \in G} \mathbf{1}_{[\tau(\pi_1(g \cdot q))=b]} \sum_{\substack{p \leq x \\ p \equiv a \pmod{k^\lambda}}} \mathbf{1}_{[T(q_0, (p)_k) = g]}$$

we find that

$$\sum_{\substack{p \leq x \\ p \equiv a \pmod{k^\lambda}}} \mathbf{1}_{[\tau(\pi_1(T(q_0, (p)_k) \cdot q))=b]} = \pi(x; 1, k^\lambda) f_{q,b} + o(\pi(x; 1, k^\lambda))$$

holds, where

$$f_{q,b} = \sum_{g \in G} f_g \mathbf{1}_{[\tau(\pi_1(g \cdot q))=b]}.$$

Therefore, we are interested in the quantity

$$\frac{1}{\varphi(k^\lambda)} \left| \{a < k^\lambda : (\varepsilon_0(a), k) = 1, \delta(q_0, (a)_k) = q\} \right|.$$

We want to show that it has a limit (for  $\lambda \rightarrow \infty$ ). We start by rewriting it as

$$\frac{1}{\varphi(k)} \sum_{\substack{a_0 < k \\ (a_0, k) = 1}} \frac{1}{k^{\lambda-1}} \left| \{a < k^{\lambda-1} : \delta(q_0, (a \cdot k + a_0)_k) = q\} \right|$$

$$= \frac{1}{\varphi(k)} \sum_{\substack{a_0 < k \\ (a_0, k) = 1}} \sum_{q' \in Q} \mathbf{1}_{[\delta(q', a_0) = q]} \frac{1}{k^{\lambda-1}} |\{a < k^{\lambda-1} : \delta(q_0, (a)_k) = q'\}|.$$

Thus it is sufficient to show that  $\frac{1}{k^{\lambda-1}} |\{a < k^{\lambda-1} : \delta(q_0, (a)_k) = q'\}|$  has a limit. However,  $\mathcal{T}$  is synchronizing and strongly connected and therefore primitive. It is a well-known result that the densities exist in this setting (see for example [1]). This allows us to write

$$|\{a < k^\lambda : (\varepsilon_0(a), k) = 1, \delta(q_0, (a)_k) = q\}| = \varphi(k^\lambda) f_q + o(\varphi(k^\lambda)),$$

where  $f_q$  does not depend on  $\lambda$ . Thus we find in total

$$\begin{aligned} \frac{1}{\pi(x)} \sum_{p \leq x} \mathbf{1}_{[a_p = b]} &= \frac{1}{\pi(x)} \sum_{q \in Q} (\varphi(k^\lambda) f_q + o(\varphi(k^\lambda))) (\pi(x; 1, k^\lambda) f_{q,b} + o(\pi(x; 1, k^\lambda))) + O(k^{-\lambda\eta}) \\ &= \sum_{q \in Q} f_q f_{q,b} + \frac{1}{\pi(x; 1, k^\lambda)} o(\pi(x; 1, k^\lambda)) + \frac{1}{\varphi(k^\lambda)} o(\varphi(k^\lambda)) + O(k^{-\lambda\eta}). \end{aligned}$$

To show that the sum of the three error terms is smaller than  $\varepsilon$ , we choose  $\lambda$  such that the second and third error term are bounded by  $\varepsilon/3$ . Then we find for  $x$  large enough that the first error term (for this given  $\lambda$ ) is also bounded by  $\varepsilon/3$  (note that  $k^{\lambda-1} \leq \varphi(k^\lambda) \leq k^\lambda$ ). Consequently we find

$$\lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} \mathbf{1}_{[a_p = b]} = \sum_{q \in Q} f_q f_{q,b},$$

which – finally – finishes the proof of Proposition 3.8. □

**Remark.** *The error terms can be made explicit. We find for example that the error term in Lemma 3.7 is actually of the form  $O(k^{-\lambda\eta'})$ . The dominant error term seems to correspond to the distribution of primes in arithmetic progressions.*

**Remark.** *As mentioned earlier, Theorem 1.3 covers block-additive functions. However, to find the corresponding result for block-additive functions the group structure degenerates to an additive structure and many arguments simplify (e.g., one does not need to use group representations).*

#### 4. A GENERAL MÖBIUS PRINCIPLE

The goal of this section is to show Proposition 3.2 and Proposition 3.6. To prove these results we use a generalization of the results in [31]. We start this section by stating the altered definitions and results. In Section 4.1 we show that for all but very special representations Definition 4.2 is fulfilled by  $f(n) = D(T(n+r))$ , uniformly in  $r$ . In Section 4.2 we show that for every representation Definition 4.1 is fulfilled by  $f(n) = D(T(n+r))$ , uniformly in  $r$ . We leave the full proofs of the generalized results of [31] to Section 5, as they are technical and very similar to the original proofs.

We consider a function  $f : \mathbb{N} \rightarrow U_d$  where  $U_d$  denotes the set of unitary  $d \times d$  matrices. Let  $k \in \mathbb{N}$ . We let  $f_\lambda$  denote the  $k^\lambda$ -periodic function defined by

$$\forall n \in \{0, \dots, k^\lambda - 1\}, \quad \forall m \in \mathbb{Z}, f_\lambda(n + mk^\lambda) = f(n).$$



Furthermore, we define  $f_{\mu,\lambda}(n) := f_\lambda(n)f_\mu(n)^H$  for  $\mu \leq \lambda$ .

It is necessary to use matrix valued functions instead of complex valued functions as in [31], as we are working with representations.

**Definition 4.1.** *A function  $f : \mathbb{N} \rightarrow U_d$  has the carry property if there exists  $\eta > 0$  such that uniformly for  $(\lambda, \alpha, \rho) \in \mathbb{N}^3$  with  $\rho < \lambda$ , the number of integers  $0 \leq \ell < k^\lambda$  such that there exists  $(n_1, n_2) \in \{0, \dots, k^\alpha - 1\}^2$  with*

$$(6) \quad f(\ell k^\alpha + n_1 + n_2)^H f(\ell k^\alpha + n_1) \neq f_{\alpha+\rho}(\ell k^\alpha + n_1 + n_2)^H f_{\alpha+\rho}(\ell k^\alpha + n_1)$$

*is at most  $O(k^{\lambda-\eta\rho})$  where the implied constant may depend only on  $k$  and  $f$ .*

**Remark.** *One can obviously exchange (6) with*

$$f(\ell k^\alpha + n_1)^H f(\ell k^\alpha + n_1 + n_2) \neq f_{\alpha+\rho}(\ell k^\alpha + n_1)^H f_{\alpha+\rho}(\ell k^\alpha + n_1 + n_2).$$

We introduce a set of functions with uniformly small Discrete Fourier Transforms as in [31]:

**Definition 4.2.** *Given a non-decreasing function  $\gamma : \mathbb{R} \rightarrow \mathbb{R}$  satisfying  $\lim_{\lambda \rightarrow \infty} \gamma(\lambda) = +\infty$  and  $c > 0$  we denote by  $F_{\gamma,c}$  the set of functions  $f : \mathbb{N} \rightarrow U_d$  such that for  $(\alpha, \lambda) \in \mathbb{N}^2$  with  $\alpha \leq c\lambda$  and  $t \in \mathbb{R}$ :*

$$(7) \quad \left\| \left\| k^{-\lambda} \sum_{u < k^\lambda} f(uk^\alpha) e(-ut) \right\|_F \right\| \leq k^{-\gamma(\lambda)}.$$

One obvious difference to [31] is that we consider matrix valued functions. Moreover, the original definition of the carry property requires  $\eta = 1$ . Nevertheless, we find results similar to those obtained in [31] in this more general setting. The results of [31] have already been generalized to matrix valued functions in [9] and to a weaker carry property (although still more restrictive than Definition 4.1) in [21]. We provide a full proof for the following theorems in Section 5.

**Theorem 4.3.** *Let  $\gamma : \mathbb{R} \rightarrow \mathbb{R}$  be a non-decreasing function satisfying  $\lim_{\lambda \rightarrow \infty} \gamma(\lambda) = +\infty$ , and  $f : \mathbb{N} \rightarrow U_d$  be a function satisfying Definition 4.1 for some  $\eta \in (0, 1]$  and  $f \in \mathcal{F}_{\gamma,c}$  for some  $c \geq 10$  in Definition 4.2. Then for any  $\theta \in \mathbb{R}$  we have*

$$(8) \quad \left\| \sum_{n \leq x} \Lambda(n) f(n) e(\theta n) \right\| \ll c_1(k) (\log x)^{c_2(k)} x k^{-\eta\gamma(2\lfloor (\log x)/(80 \log k) \rfloor / 20)},$$

*with the same constants as in [31].*

**Theorem 4.4.** *Let  $\gamma : \mathbb{R} \rightarrow \mathbb{R}$  be a non-decreasing function satisfying  $\lim_{\lambda \rightarrow \infty} \gamma(\lambda) = +\infty$ , and  $f : \mathbb{N} \rightarrow U_d$  be a function satisfying Definition 4.1 for some  $\eta \in (0, 1]$  and  $f \in \mathcal{F}_{\gamma,c}$  for some  $c \geq 10$  in Definition 4.2. Then for any  $\theta \in \mathbb{R}$  we have*

$$(9) \quad \left\| \sum_{n \leq x} \mu(n) f(n) e(\theta n) \right\| \ll c_1(k) (\log x)^{\frac{9}{4} + \frac{1}{4} \max(\omega(k), 2)} x k^{-\eta\gamma(2\lfloor (\log x)/(80 \log k) \rfloor / 20)},$$

*with the same constants as in [31].*

**4.1. Fourier Estimates.** One of the most difficult parts of this approach is to find sharp enough bounds for the Fourier terms. The goal of this subsection is to show that Definition 4.2 holds for all but very special unitary irreducible representations where  $f(n) = D(T(n+r))$  and  $\gamma(\lambda) = \eta'\lambda - c'$  for some  $\eta' > 0$  and  $c' \in \mathbb{R}$ . The proof is rather technical. However, it justifies some results of Section 2.

Let  $\mathcal{T}_A$  be a naturally induced transducer of  $A$  and suppose that  $d(A) = 1, k_0(A) = 1$  holds. We remind the reader of the following special representations.

**Lemma 3.5.** *There exist  $d'$  representations  $D_0, \dots, D_{d'-1}$  defined by*

$$D_\ell(g) := e\left(\frac{\ell \cdot s_0(g)}{d'}\right),$$

for  $\ell = 0, \dots, d' - 1$ , where  $s_0$  is defined by  $s_0(T(q, \mathbf{w})) \equiv [\mathbf{w}]_k \pmod{d'}$ .

Note that  $D_\ell(T(q_0, \mathbf{w})) = e((\ell \cdot [\mathbf{w}]_k)/d')$  for  $\ell = 0, \dots, d' - 1, \mathbf{w} \in \Sigma^*$ .

We try to find exponentially decreasing bounds for expressions similar to

$$\left\| \frac{1}{k^\lambda} \sum_{u < k^\lambda} D(T(q_0, (u)_k)) e(-ut) \right\|_2,$$

uniformly in  $t$ , where  $D$  is a unitary, irreducible representation of  $G$ .

We already see that  $D = D_\ell$  is a special case for which we are not able to find exponentially decreasing bounds as

$$\frac{1}{k^\lambda} \sum_{u < k^\lambda} D_\ell(T(q_0, (u)_k)) e(-ut) = \frac{1}{k^\lambda} \sum_{u < k^\lambda} e\left(u \left(\frac{\ell}{d'} - t\right)\right)$$

gives 1 for  $t = \ell/d'$ , which is the trivial bound. Nevertheless, we are able to find exponentially decreasing bounds for all other unitary and irreducible representations of  $G$ :

**Theorem 4.5.** *Let  $D$  be a unitary and irreducible representation of  $G$  different from  $D_0, \dots, D_{d'-1}$ . Then there exists  $\eta > 0$  such that*

$$(10) \quad \left\| \frac{1}{k^\lambda} \sum_{u < k^\lambda} D(T(q, (uk^\alpha + r)_k)) e(-ut) \right\|_2 \ll k^{-\eta\lambda}$$

holds uniformly for  $t \in \mathbb{R}, q \in Q, r \in \mathbb{N}$  and  $\alpha \in \mathbb{N}$ .

The proof is carried out throughout this subsection. We define

$$\phi_{\lambda, \alpha}^q(t, r) := \frac{1}{k^\lambda} \sum_{u < k^\lambda} D(T(q, (uk^\alpha + r)_k)) e(-ut)$$

and for  $r < k^\alpha$

$$\psi_{\lambda, \alpha}^q(t, r) := \frac{1}{k^\lambda} \sum_{u < k^\lambda} D(T(q, (u)_k^\lambda (r)_k^\alpha)) e(-ut).$$

We see that these two definitions look very similar, but it turns out to be much easier to deal with  $\psi$ .

**Lemma 4.6.** *Let  $D$  be a unitary representation of  $G$ . Assume that there exists  $\eta > 0$  such that*

$$\|\psi_{\lambda,\alpha}^q(t,r)\|_2 \ll k^{-\eta\lambda}$$

*holds uniformly for  $q \in Q, \lambda, \alpha \in \mathbb{N}, t \in \mathbb{R}$  and  $r < k^\alpha$ . Then Theorem 4.5 holds for  $D$ .*

*Proof.* We postpone this proof to Section 5. □

Therefore, it is sufficient to prove the following proposition.

**Proposition 4.7.** *Let  $D$  be a unitary and irreducible representation of  $G$  different from  $D_0, \dots, D_{d-1}$ . Then there exists  $\eta > 0$  such that*

$$\frac{1}{k^\lambda} \left\| \sum_{u < k^\lambda} D(T(q, (u)_k^\lambda (r)_k^\alpha)) e(-ut) \right\|_2 \ll k^{-\eta\lambda}$$

*holds uniformly for  $q \in Q, \lambda, \alpha \in \mathbb{N}, t \in \mathbb{R}$  and  $r < k^\alpha$ .*

We use the recursive structure of transducers to find recurrences for these Fourier terms.

**Lemma 4.8.** *Let  $q \in Q, \lambda, \alpha \in \mathbb{N}, t \in \mathbb{R}$  and  $r < k^\alpha$ . It holds for all  $m < \lambda$*

$$(11) \quad \psi_{\lambda,\alpha}^q(t,r) = \frac{1}{k^m} \sum_{\varepsilon < k^m} D(T(q, (\varepsilon)_k^m)) e(-\varepsilon(k^{\lambda-m}t)) \psi_{\lambda-m,\alpha}^{\delta(q,\varepsilon)}(t,r).$$

*Proof.* We know that  $T(q, (\varepsilon)_k^m (u')_k^{\lambda-m} (r)_k^\alpha) = T(q, (\varepsilon)_k^m) \cdot T(\delta(q, (\varepsilon)_k^m), (u')_k^{\lambda-m} (r)_k^\alpha)$  for  $u' < k^{\lambda-m}$ . By distinguishing the  $m$  most significant digits of  $u$ , we find

$$\begin{aligned} \psi_{\lambda,\alpha}^q(t) &= \frac{1}{k^\lambda} \sum_{\varepsilon < k^m} \sum_{u' < k^{\lambda-m}} D(T(q, (\varepsilon)_k^m (u')_k^{\lambda-m} (r)_k^\alpha)) e(-(\varepsilon k^{\lambda-m} + u')t) \\ &= \frac{1}{k^m} \sum_{\varepsilon < k^m} D(T(q, (\varepsilon)_k^m)) e(-\varepsilon(k^{\lambda-m}t)) \\ &\quad \frac{1}{k^{\lambda-m}} \sum_{u' < k^{\lambda-m}} D(T(\delta(q, (\varepsilon)_k^m), (u')_k^{\lambda-m} (r)_k^\alpha)) e(-u't). \end{aligned}$$

□

The main idea is to find  $m \in \mathbb{N}, \varepsilon_1, \varepsilon_2 < k^m$  such that two or more terms on the right side of (11) cancel – at least partially. This means that we want to find  $\varepsilon_1, \varepsilon_2$  such that  $\delta(q, \varepsilon_1) = \delta(q, \varepsilon_2)$  and

$$\|D(T(q, \varepsilon_1)) e(-\varepsilon_1 k^{\lambda-m}t) + D(T(q, \varepsilon_2)) e(-\varepsilon_2 k^{\lambda-m}t)\|_2 < 2.$$

We split the proof into two parts – depending on the dimension of  $D$ .

**Lemma 4.9.** *Let  $D$  be a unitary and irreducible representation of  $G$  of dimension at least 2. Then Proposition 4.7 holds for  $D$ .*

*Proof.* We need to show that there exists  $\eta > 0$  such that Proposition 4.7 holds. Theorem 2.7 shows that there exists for every  $g \in G$  a path  $\mathbf{w}_g \in \Sigma^{m_0}$  such that  $\delta(q, \mathbf{w}_g) = q, T(q, \mathbf{w}_g) = g$  holds. We find using Lemma 4.8 that

$$\begin{aligned}
\|\psi_{\lambda, \alpha}^q(t, r)\|_2 &= \frac{1}{k^{m_0}} \left\| \sum_{\varepsilon < k^{m_0}} D(T(q, (\varepsilon)_k^{m_0})) e(-\varepsilon(k^{\lambda-m_0}t)) \psi_{\lambda-m_0, \alpha}^{\delta(q, (\varepsilon)_k^{m_0})}(t, r) \right\|_2 \\
&\leq \frac{1}{k^{m_0}} \left\| \sum_{g \in G} D(T(q, \mathbf{w}_g)) e(-[\mathbf{w}_g]_k k^{\lambda-m_0}t) \psi_{\lambda-m_0, \alpha}^q(t, r) \right\|_2 \\
&\quad + \frac{1}{k^{m_0}} \sum_{\substack{\varepsilon < k^{m_0} \\ \varepsilon \neq [\mathbf{w}_g]_k \forall g \in G}} \left\| D(T(q, (\varepsilon)_k^{m_0})) e(-\varepsilon(k^{\lambda-m_0}t)) \psi_{\lambda-m_0, \alpha}^{\delta(q, (\varepsilon)_k^{m_0})}(t, r) \right\|_2 \\
&\leq \frac{1}{k^{m_0}} \left\| \sum_{g \in G} D(T(q, \mathbf{w}_g)) e(-[\mathbf{w}_g]_k k^{\lambda-m_0}t) \right\|_2 \cdot \|\psi_{\lambda-m_0, \alpha}^q(t, r)\|_2 \\
&\quad + \frac{k^{m_0} - |G|}{k^{m_0}} \max_{q' \in Q} \|\psi_{\lambda-m_0, \alpha}^{q'}(t, r)\|_2 \\
&\leq \frac{1}{k^{m_0}} \left( k^{m_0} - |G| + \left\| \sum_{g \in G} D(g) e(-[\mathbf{w}_g]_k k^{\lambda-m_0}t) \right\|_2 \right) \max_{q' \in Q} \|\psi_{\lambda-m_0, \alpha}^{q'}(t, r)\|_2.
\end{aligned}$$

Thus it is sufficient to show that for all  $t' \in \mathbb{R}$  holds

$$(12) \quad \left\| \sum_{g \in G} D(g) e(-[\mathbf{w}_g]_k t') \right\|_2 < |G|.$$

Since the left side is a periodic and continuous function in  $t'$ , this implies that there exists  $\eta' > 0$  such that

$$\left\| \sum_{g \in G} D(g) e(-[\mathbf{w}_g]_k t') \right\|_2 \leq |G| - \eta'$$

for all  $t' \in \mathbb{R}$ . This gives in total

$$\|\psi_{\lambda, \alpha}^q(t, r)\|_2 \leq \left(1 - \frac{\eta'}{k^{m_0}}\right) \max_{q' \in Q} \|\psi_{\lambda-m_0, \alpha}^{q'}(t, r)\|_2$$

and the statement follows easily as  $\eta'$  only depends on  $G$  and  $D$  (but not on  $\lambda, \alpha, t'$  or  $r$ ). What follows is a variation of the proof of [10, Lemma 4]:

Let us assume – on the contrary – that there exists  $t' \in \mathbb{R}$  such that

$$\left\| \sum_{g \in G} D(g) e(-[\mathbf{w}_g]_k t') \right\|_2 = |G|.$$

This holds if and only if there exists  $\mathbf{0} \neq \mathbf{y} \in \mathbb{C}^d$  such that

$$\left\| \sum_{g \in G} D(g) e(-[\mathbf{w}_g]_k t') \mathbf{y} \right\|_2^2 = \sum_{g_1, g_2 \in G} \langle D(g_1) e(-[\mathbf{w}_{g_1}]_k t') \mathbf{y}, D(g_2) e(-[\mathbf{w}_{g_2}]_k t') \mathbf{y} \rangle = |G|^2 \|\mathbf{y}\|_2^2$$

However, the Cauchy-Schwarz inequality implies

$$(13) \quad \begin{aligned} & |\langle D(g_1) e(-[\mathbf{w}_{g_1}]_k t') \mathbf{y}, D(g_2) e(-[\mathbf{w}_{g_2}]_k t') \mathbf{y} \rangle| \\ & \leq \|D(g_1) e(-[\mathbf{w}_{g_1}]_k t') \mathbf{y}\|_2^2 \|D(g_2) e(-[\mathbf{w}_{g_2}]_k t') \mathbf{y}\|_2^2 = \|\mathbf{y}\|_2^2. \end{aligned}$$

For equality to hold in (13) it is necessary that the  $D(g_i) e(-[\mathbf{w}_{g_i}]_k t')$  are linearly dependent. Since we find for  $g_1 = g_2 = id$  the summand  $\langle e(-[\mathbf{w}_{id}]_k t') \mathbf{y}, e(-[\mathbf{w}_{id}]_k t') \mathbf{y} \rangle$  we obtain for all  $g \in G$

$$D(g) e(-[\mathbf{w}_g]_k t') \mathbf{y} = e(-[\mathbf{w}_{id}]_k t') \mathbf{y},$$

i.e.,  $\mathbf{y}$  is an eigenvector of all  $D(g)$ ,  $g \in G$ . We define  $W = \text{span}(\mathbf{y})$  and find  $D(g)W \subseteq W$  for all  $g \in G$ . This means that  $D$  would be reducible which yields a contradiction.

Therefore (12) holds, which concludes this proof.  $\square$

**Lemma 4.10.** *Let  $D$  be a one-dimensional, unitary and irreducible representation of  $G$  different from  $D_0, \dots, D_{d'-1}$ . Then Proposition 4.7 holds for  $D$ .*

*Proof.* Our goal is to show that there exists some  $\eta' > 0$  (only depending on  $G$  and  $D$ ) such that

$$(14) \quad \|\psi_{\lambda, \alpha}^q(t, r)\|_2 \leq (1 - \eta') \max_{\tilde{q} \in Q} \left\| \psi_{\lambda - m'_0 - \ell_0, \alpha}^{\tilde{q}}(t, r) \right\|_2$$

holds for all  $t \in \mathbb{R}$ . This implies again Proposition 4.7 for some  $\eta > 0$  as in the proof of the previous case.

We need two different estimates for this step and start to work on the first estimate.

We start by using the properties of  $G$ , i.e.,  $\{g \in G : s_0(g) = \ell\} = G_0 \cdot (g'_0)^\ell$ , to find restrictions for  $D$ .

Assume that  $D(g) = 1$  for all  $g \in G_0$ . It follows that  $D(g'_0) = e(\ell/d')$  for some  $\ell < d'$  since  $(g'_0)^{d'} \in G_0$  and therefore we see directly that  $D = D_\ell$ . Thus we know that there exists  $g \in G_0$  such that  $D(g) \neq 1$ .

We use Lemma 2.13 to see that there exist  $\mathbf{w}_{id}, \mathbf{w}_g \in \Sigma^{m'_0}$  such that  $\delta(q, \mathbf{w}_{id}) = \delta(q, \mathbf{w}_g) = q$  and  $T(q, \mathbf{w}_{id}) = id, T(q, \mathbf{w}_g) = g$  holds. By using Lemma 4.8 (with  $m = m'_0$ ), we find (as in the proof for representations of dimension  $\geq 2$ )

$$\begin{aligned} \|\psi_{\lambda, \alpha}^q(t, r)\|_2 & \leq \frac{1}{k^{m'_0}} \left\| D(T(q, \mathbf{w}_{id})) e(-[\mathbf{w}_{id}]_k k^{\lambda - m'_0} t) \psi_{\lambda - m'_0, \alpha}^{\delta(q, \mathbf{w}_{id})}(t, r) \right. \\ & \quad \left. + D(T(q, \mathbf{w}_g)) e(-[\mathbf{w}_g]_k k^{\lambda - m'_0} t) \psi_{\lambda - m'_0, \alpha}^{\delta(q, \mathbf{w}_g)}(t, r) \right\|_2 \\ & \quad + \frac{k^{m'_0} - 2}{k^{m'_0}} \max_{q' \in Q} \left\| \psi_{\lambda - m'_0, \alpha}^{q'}(t, r) \right\|_2. \end{aligned}$$

We see that the first term of the right side equals

$$\begin{aligned} & \frac{1}{k^{m'_0}} \left\| 1 \cdot e(-[\mathbf{w}_{id}]_k k^{\lambda - m'_0} t) + D(g) \cdot e(-[\mathbf{w}_g]_k k^{\lambda - m'_0} t) \right\|_2 \left\| \psi_{\lambda - m'_0, \alpha}^q(t, r) \right\|_2 \\ & = \frac{1}{k^{m'_0}} \left\| 1 + D(g) e(-([\mathbf{w}_g]_k - [\mathbf{w}_{id}]_k) k^{\lambda - m'_0} t) \right\|_2 \left\| \psi_{\lambda - m'_0, \alpha}^q(t, r) \right\|_2. \end{aligned}$$

Now we use Lemma 4.8 again with  $m = \ell_0$  to find in total

$$(15) \quad \left\| \psi_{\lambda, \alpha}^q(t, r) \right\|_2 \leq \frac{k^{m'_0} - 2 + \left| 1 + D(g) e(-([\mathbf{w}_g]_k - [\mathbf{w}_{id}]_k) k^{\lambda - m'_0} t) \right|}{k^{m'_0}} \max_{\tilde{q} \in Q} \left\| \psi_{\lambda - \ell_0 - m'_0, \alpha}^{\tilde{q}}(t, r) \right\|_2.$$

To find the second estimate, we use Lemma 4.8 with  $m = \ell_0$  and find

$$\left\| \psi_{\lambda, \alpha}^q(t, r) \right\|_2 \leq \max_{\tilde{q} \in Q} \left\| \psi_{\lambda - \ell_0, \alpha}^{\tilde{q}}(t, r) \right\|_2.$$

For convenience we assume that the maximal value of the right side is attained at  $q$ . Theorem 2.16 shows  $d' \cdot d''(q, q) = \gcd\{[\mathbf{w}_1]_k - [\mathbf{w}_2]_k : \mathbf{w}_i \in \Sigma^{m'_0} \text{ with } \delta(q, \mathbf{w}_i) = q, T(q, \mathbf{w}_i) = id\}$ . Therefore, there exist  $N \in \mathbb{N}$ ,  $\mu_{i,j} \in \mathbb{Z}^N \times \mathbb{Z}^N$  such that  $d' \cdot d''(q, q) = \sum_{i,j < N} \mu_{i,j} ([\mathbf{w}_i]_k - [\mathbf{w}_j]_k)$  where  $\mathbf{w}_i, \mathbf{w}_j \in \Sigma^{m'_0}$  such that  $\delta(q, \mathbf{w}_i) = \delta(q, \mathbf{w}_j) = q$  and  $T(q, \mathbf{w}_i) = T(q, \mathbf{w}_j) = id$ . By Lemma 4.8 and the same arguments we used above, we find

$$\begin{aligned} \left\| \psi_{\lambda - \ell_0, \alpha}^q(t, r) \right\|_2 &\leq \frac{1}{k^{m'_0}} \left\| \sum_{i < N} D(T(q, \mathbf{w}_i)) e(-[\mathbf{w}_i]_k k^{\lambda - \ell_0 - m'_0} t) \psi_{\lambda - \ell_0 - m'_0, \alpha}^q(t, r) \right\|_2 \\ &\quad + \frac{k^{m'_0} - N}{k^{m'_0}} \max_{\tilde{q} \in Q} \left\| \psi_{\lambda - \ell_0 - m'_0, \alpha}^{\tilde{q}}(t, r) \right\|_2 \\ &= \frac{1}{k^{m'_0}} \left\| \sum_{i < N} e(-[\mathbf{w}_i]_k k^{\lambda - \ell_0 - m'_0} t) \right\| \left\| \psi_{\lambda - \ell_0 - m'_0, \alpha}^q(t, r) \right\|_2 \\ &\quad + \frac{k^{m'_0} - N}{k^{m'_0}} \max_{\tilde{q} \in Q} \left\| \psi_{\lambda - \ell_0 - m'_0, \alpha}^{\tilde{q}}(t, r) \right\|_2. \end{aligned}$$

Thus, we find in total

$$(16) \quad \left\| \psi_{\lambda, \alpha}^q(t, r) \right\|_2 \leq \frac{k^{m'_0} - N + \left| \sum_{i < N} e(-[\mathbf{w}_i]_k k^{\lambda - \ell_0 - m'_0} t) \right|}{k^{m'_0}} \max_{\tilde{q} \in Q} \left\| \psi_{\lambda - \ell_0 - m'_0, \alpha}^{\tilde{q}}(t, r) \right\|_2.$$

We combine (15) and (16) and find

$$\begin{aligned} \left\| \psi_{\lambda, \alpha}^q(t, r) \right\|_2 &\leq \left( 1 - \frac{2 - \left| 1 + D(g) e(-([\mathbf{w}_g]_k - [\mathbf{w}_{id}]_k) k^{\lambda - m'_0} t) \right| + N - \left| \sum_{i < N} e(-[\mathbf{w}_i]_k k^{\lambda - \ell_0 - m'_0} t) \right|}{2k^{m'_0}} \right) \\ &\quad \max_{\tilde{q} \in Q} \left\| \psi_{\lambda - \ell_0 - m'_0, \alpha}^{\tilde{q}}(t, r) \right\|_2. \end{aligned}$$

It remains to show that

$$(17) \quad \left| 1 + D(g) e(-([\mathbf{w}_g]_k - [\mathbf{w}_{id}]_k) k^{\lambda - m'_0} t) \right| + \left| \sum_{i < N} e(-[\mathbf{w}_i]_k k^{\lambda - \ell_0 - m'_0} t) \right| < N + 2$$

holds for all  $t \in \mathbb{R}$  as the left side is a periodic and continuous function in  $t$ . We distinguish the following two cases.

At first let us assume  $d'tk^{\lambda - m'_0} \in \mathbb{Z}$ :

Since  $g \in G_0$  we know that  $[\mathbf{w}_{id}] \equiv [\mathbf{w}_g] \pmod{d'}$  and, therefore, the first term of the left side of equation (17) simplifies to  $|1 + D(g)|$ . By the definition of  $g$ , we find that  $|1 + D(g)| < 2$  and, therefore, equation (17) holds.

The remaining case is  $d'tk^{\lambda-m'_0} \notin \mathbb{Z}$ :

Let us assume

$$\left| \sum_{i < N} e(-[\mathbf{w}_i]_k k^{\lambda-\ell_0-m'_0} t) \right| = N.$$

This implies  $([\mathbf{w}_i]_k - [\mathbf{w}_j]_k) k^{\lambda-\ell_0-m'_0} t \in \mathbb{Z}$  for all  $i, j < N$ . As  $d' \cdot d''(q_0, q_0)$  is a linear combination of  $([\mathbf{w}_i]_k - [\mathbf{w}_j]_k)$ , we find  $d' \cdot d''(q, q) k^{\lambda-\ell_0-m'_0} t \in \mathbb{Z}$ . This yields a contradiction since  $d''(q, q) | k^{\ell_0}$ . Therefore, we have

$$\left| \sum_{i < N} e(-[\mathbf{w}_i]_k k^{\lambda-\ell_0-m'_0} t) \right| < N$$

and, consequently, equation (17) holds, which finishes the proof.  $\square$

**4.2. Carry Lemma.** We show in this subsection that the function  $f(n) = D(T(q_0, (n+r)_k))$  has the carry property – uniformly in  $r$ .

**Lemma 4.11.** *Definition 4.1 holds – uniformly in  $r$  – for  $f(n) = D(T(q_0, (n+r)_k))$  where  $D$  is a unitary and irreducible representation of  $G$ ,  $\eta$  is given by [7, Lemma 2.2], i.e.,  $\eta = \frac{\log(k^{\ell_0}/(k^{\ell_0}-1))}{\log(k^{\ell_0})}$ , and the implied constant does not depend on  $r$ .*

*Proof.* We fix  $\lambda, \rho$  and  $\alpha$ . We rewrite  $r = r_1 k^\alpha + r_2$  where  $r_1, r_2 \in \mathbb{N}, r_2 < k^\alpha$ . We want to distinguish the form of  $\ell + r_1$ , i.e., the maximal number  $t$  such that the digits at position  $\alpha, \dots, \alpha + t$  of  $(\ell + r_1)k^\alpha + n_1 + n_2 + r_2$  can be affected by the carry of  $n_1 + n_2 + r_2$ . As  $n_1 + n_2 + r_2 \leq 3k^\alpha - 3$  limits the carry to 0, 1, or 2, we define  $t := \max(\nu_k(\ell + r_1 + 1), \nu_k(\ell + r_1 + 2))$ . Thus we know that the digits at position  $\alpha + t + i$  of  $(\ell k^\alpha + n_1 + n_2 + r)$  and  $(\ell k^\alpha + n_1 + r)$  coincide for  $i = 1, 2, \dots$  and are equal to the digits at position  $t + i$  of  $\ell + r_1$ . We fix  $t$  such that  $t < \rho$  and count the number of integers  $\ell$  such that (6) holds for some  $n_1, n_2$ . As the terms corresponding to the digits of index  $\alpha + t + i$  cancel for  $i \geq 1$ , we find

$$\begin{aligned} & D(T(q_0, (\ell k^\alpha + n_1 + n_2 + r)_k))^H D(T(q_0, (\ell k^\alpha + n_1 + r)_k)) \\ &= D(T(\delta(q_0, (\lfloor (\ell + r_1)/k^t \rfloor)_k), (\ell k^\alpha + n_1 + n_2 + r)_k^{\alpha+t}))^H \\ & \quad D(T(\delta(q_0, (\lfloor (\ell + r_1)/k^t \rfloor)_k), (\ell k^\alpha + n_1 + r)_k^{\alpha+t})) \end{aligned}$$

and

$$\begin{aligned} & D(T_{\alpha+\rho}(q_0, (\ell k^\alpha + n_1 + n_2 + r)_k))^H D(T_{\alpha+\rho}(q_0, (\ell k^\alpha + n_1 + r)_k)) \\ &= D(T(\delta(q_0, (\lfloor (\ell + r_1)/k^t \rfloor)_k^{\rho-t}), (\ell k^\alpha + n_1 + n_2 + r)_k^{\alpha+t}))^H \\ & \quad D(T(\delta(q_0, (\lfloor (\ell + r_1)/k^t \rfloor)_k^{\rho-t}), (\ell k^\alpha + n_1 + r)_k^{\alpha+t})). \end{aligned}$$

Thus, the number of integers  $\ell$  (with fixed  $t$  and  $r$ ) for which (6) holds for some  $n_1, n_2$  can be bounded by the number of integers  $\ell$  such that  $(\lfloor (\ell + r_1)/k^t \rfloor)_k^{\rho-t}$  is not synchronizing – otherwise  $\delta(q_0, (\lfloor (\ell + r_1)/k^t \rfloor)_k) = \delta(q_0, (\lfloor (\ell + r_1)/k^t \rfloor)_k^{\rho-t})$  would imply that (6) does not hold. By [7, Lemma 2.2], we know that this number is bounded by  $O(k^{(\rho-t)(1-\eta)} k^{\lambda-\rho})$  for some  $\eta > 0$  where the constant only depends on  $k$  and  $A$ . Note that there are only two

possibilities for the digits with index  $0, \dots, t-1$ . By summing over  $t$  we find that the number of such  $\ell$  is bounded by

$$\begin{aligned} c \sum_{t \leq \rho} k^{\lambda-\rho} k^{(\rho-t)(1-\eta)} + k^{\lambda-\rho} &\leq ck^{\lambda-\eta\rho} \sum_{t \leq \rho} k^{-t(1-\eta)} + k^{\lambda-\rho} \\ &\leq ck^{\lambda-\eta\rho} \frac{1}{1-k^{-(1-\eta)}} + k^{\lambda-\rho} \end{aligned}$$

and the result follows easily as  $c$  does not depend on  $\lambda, \rho, \alpha$  and  $r$ .  $\square$

**4.3. Proof of Proposition 3.2 and Proposition 3.6.** We are now ready to show Proposition 3.2 and, thereafter, Proposition 3.6.

The proof for Proposition 3.2 differs when  $D = D_\ell$  and we start by considering this specific case.

**Lemma 4.12.** *For all  $\ell < d'$ ,  $\lambda_1, \lambda_2, r \in \mathbb{N}$  and  $b < k^{\lambda_1}, m < k^{\lambda_2}$  we have*

$$\left\| \sum_{\substack{n < N \\ n \equiv m \pmod{k^{\lambda_2}}} \chi_{k^{-\lambda_1}} \left( \frac{n+r-bk^{\nu-\lambda_1}}{k^\nu} \right) D_\ell(T(q_0, (n+r)_k)) \mu(n) \right\|_F = o(N)$$

uniformly in  $r \in \mathbb{N}$ .

*Proof.* We find by the characterizing property of  $D_\ell$  (i.e.,  $D_\ell(T(q_0, (n)_k)) = e(\ell n/d')$ )

$$\begin{aligned} &\left\| \sum_{\substack{n < N \\ n \equiv m \pmod{k^{\lambda_2}}} \chi_{k^{-\lambda_1}} \left( \frac{n+r-bk^{\nu-\lambda_1}}{k^\nu} \right) D_\ell(T(q_0, (n+r)_k)) \mu(n) \right\|_F \\ &= \left| \sum_{\substack{n < N \\ n \equiv m \pmod{k^{\lambda_2}}} \chi_{k^{-\lambda_1}} \left( \frac{n+r-bk^{\nu-\lambda_1}}{k^\nu} \right) e \left( (n+r) \frac{\ell}{d'} \right) \mu(n) \right| \\ &\leq \sum_{\substack{m' < k^{\lambda_2} d' \\ m' \equiv m \pmod{k^{\lambda_2}}} \left| e \left( m' \frac{\ell}{d'} \right) \right| \left| \sum_{n < N} \chi_{k^{-\lambda_1}} \left( \frac{n+r-bk^{\nu-\lambda_1}}{k^\nu} \right) \mathbf{1}_{[n+r \equiv m' \pmod{k^{\lambda_2} d'}]} \mu(n) \right|. \end{aligned}$$

As

$$\sum_{n < N} \chi_{k^{-\lambda_1}} \left( \frac{n+r-bk^{\nu-\lambda_1}}{k^\nu} \right)$$

is indeed a sum over at most two intervals included in  $[1, \dots, N-1]$ , the result follows from the well-known result

$$\sum_{\substack{n < N \\ n \equiv r \pmod{s}}} \mu(n) = o(N).$$

$\square$



*Proof of Proposition 3.2:* For  $D = D_\ell$  Lemma 4.12 gives the desired result. Suppose from now on  $D \neq D_\ell$  for all  $\ell < d'$ . We rewrite the left side of (3) using exponential sums and obtain using Vaaler-approximation (see for example [31, Lemma 1])

$$\begin{aligned}
& \left\| \sum_{n < N} \frac{1}{k^{\lambda_2}} \sum_{h < k^{\lambda_2}} e\left(h \frac{n-m}{k^{\lambda_2}}\right) \sum_{n < N} \chi_{k^{-\lambda_1}}\left(\frac{n+r-bk^{\nu-\lambda_1}}{k^\nu}\right) D(T(n+r))\mu(n) \right\|_F \\
& \leq \frac{1}{k^{\lambda_2}} \sum_{h < k^{\lambda_2}} \left\| \sum_{n < N} e\left(n \frac{h}{k^{\lambda_2}}\right) A_{k^{-\lambda_1}, H}\left(\frac{n+r-bk^{\nu-\lambda_1}}{k^\nu}\right) D(T(n+r))\mu(n) \right\|_F \\
& \quad + \frac{1}{k^{\lambda_2}} \sum_{h < k^{\lambda_2}} \sum_{n < N} B_{k^{-\lambda_1}, H}\left(\frac{n+r-bk^{\nu-\lambda_1}}{k^\nu}\right) \left\| e\left(n \frac{h}{k^{\lambda_2}}\right) D(T(n+r))\mu(n) \right\|_F \\
& \ll \frac{1}{k^{\lambda_2}} \sum_{h < k^{\lambda_2}} \left\| \sum_{n < N} e\left(n \frac{h}{k^{\lambda_2}}\right) \sum_{|h'| \leq H} a_{h'}(k^{-\lambda_1}, H) e\left(h' \frac{n+r-bk^{\nu-\lambda_1}}{k^\nu}\right) D(T(n+r))\mu(n) \right\|_F \\
& \quad + \sum_{n < N} \sum_{|h'| \leq H} b_{h'}(k^{-\lambda_1}, H) e\left(h' \frac{n+r-bk^{\nu-\lambda_1}}{k^\nu}\right),
\end{aligned}$$

where the implied constant only depends on the dimension of  $D$  and  $|a_{h'}(k^{-\lambda_1}, H)| \leq \min\left(k^{-\lambda_1}, \frac{1}{\pi|h'}\right)$ ,  $|b_{h'}(k^{-\lambda_1}, H)| \leq \frac{1}{H+1}$  holds. We assume that  $k^{\lambda_1} \leq H \leq k^\nu$ .

Thus, the first summand above is bounded by

$$\begin{aligned}
& \frac{1}{k^{\lambda_2}} \sum_{h < k^{\lambda_2}} \sum_{|h'| < H} |a_{h'}(k^{-\lambda_1}, H)| \left\| \sum_{n < N} e\left(n \left(\frac{h}{k^{\lambda_2}} + \frac{h'}{k^\nu}\right)\right) D(T(n+r))\mu(n) \right\|_F \\
& \ll \left| \sum_{h' < k^{\lambda_1}} k^{-\lambda_1} + \sum_{k^{\lambda_1} < h' < H} \frac{1}{h'} \right| \sup_{\theta \in \mathbb{R}} \left\| \sum_{n < N} e(\theta n) D(T(n+r))\mu(n) \right\|_F \\
& \ll (1 + \log(H)) \sup_{\theta \in \mathbb{R}} \left\| \sum_{n < N} e(\theta n) D(T(n+r))\mu(n) \right\|_F
\end{aligned}$$

However, as  $D(T(n+r))$  fulfills Definition 4.1 and Definition 4.2 uniformly in  $r$ , we can apply Theorem 4.4. This gives for some  $\eta' > 0$  the asymptotic estimate  $(1 + \log(H))N^{1-\eta'}$ .

The second summand is positive and can be estimated by classical estimates involving geometric series (see for example [31]).

$$\begin{aligned}
& \sum_{|h'| \leq H} b_{h'}(k^{-\lambda_1}, H) \sum_{n < N} e\left(h' \frac{n+r-bk^{\nu-\lambda_1}}{k^\nu}\right) \\
& \leq \sum_{|h'| \leq H} \frac{1}{H} \left| \sum_{n < N} e\left(h' \frac{n}{k^\nu}\right) \right| \\
& \leq \frac{1}{H} \sum_{|h'| \leq k^\nu} \min(N, |\sin(\pi h' k^{-\nu})|)
\end{aligned}$$

$$\begin{aligned} &\ll \frac{1}{H}(N + k^\nu \log(k^\nu)) \\ &\ll \frac{N \log(N)}{H}, \end{aligned}$$

where the constant is absolute.

Thus choosing for example  $H = \lfloor N^{1/2} \rfloor$  finishes the proof.  $\square$

*Proof of Proposition 3.6:* By classical partial summation one finds the following well-known result (see for example [30, Lemma 11]).

If  $f : \mathbb{N} \rightarrow \mathbb{C}$  is such that  $|f(n)| \leq 1$  for all  $n \in \mathbb{N}$  then

$$(18) \quad \left| \sum_{p \leq x} f(p) \right| \leq \frac{2}{\log x} \max_{t \leq x} \left| \sum_{n \leq t} \Lambda(n) f(n) \right| + O(\sqrt{x}).$$

This result can easily be adapted to matrix valued functions. Thus we find

$$\left\| \frac{1}{\pi(k^\nu)} \sum_{p < k^\nu} D(T(q_0, (p)_k)) e(pt) \right\|_F \ll \frac{1}{\log k^\nu} \frac{1}{\pi(k^\nu)} \max_{t \leq k^\nu} \left\| \sum_{n \leq t} \Lambda(n) D(T(q_0, (n)_k)) e(nt) \right\|_F + O(k^{-\nu/2})$$

As  $D(T(n))$  fulfills Definition 4.1 and Definition 4.2, we can apply Theorem 4.3 and find that the right side is asymptotically bounded by

$$\max_{t \leq k^\nu} t^{-\eta'} + O(k^{-\nu/2})$$

for some  $\eta' > 0$  – if we only choose  $\eta'$  strictly smaller than  $\eta$  to remove the other terms. This gives directly the desired result.  $\square$

## 5. TECHNICAL RESULTS

In this section, we will provide the proofs for some results mentioned earlier, where the proof is technical and seemed to disturb the flow of reading.

**5.1. Number of words with special properties.** The first result stated that some properties occur “often” in the digital representation of numbers. We considered an automaton with final components  $Q'_i$ . Furthermore, we let  $\mathbf{w}_i$  denote the synchronizing word for the naturally induced transducer corresponding to the automaton that is the restriction of the original automaton to  $Q'_i$ .

**Lemma 3.3.** *The set*

$M := \{n \in \mathbb{N} \mid \exists \mathbf{v}_1, \mathbf{v}_2 \in \Sigma^* : (n)_k = \mathbf{v}_1 \mathbf{v}_2 \wedge \forall q' \in Q' : \delta'(q', \mathbf{v}_1) \in \cup Q'_i \wedge \forall i : \mathbf{w}_i \text{ is a subword of } \mathbf{v}_2\}$   
has density 1.

*Proof.* One sees easily that it suffices to show that

$$\lim_{\nu \rightarrow \infty} \frac{M \cap [0, k^{2\nu})}{k^{2\nu}} = 1$$

holds. Without loss of generality we assume that  $\mathbf{w}_1$  has no leading zeros. We define  $M_1 := \{n \in \mathbb{N} \mid \forall q' \in Q' : \delta'(q', (n)_k) \in \cup Q'_i\}$ ,  $M_{2,\nu} := \{n < k^\nu \mid \mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_\ell \text{ is a subword of } (n)_k^\nu\}$  and  $M_2 := \lim_{\nu \rightarrow \infty} M_{2,\nu}$ . The idea is to show that both sets have density one.

Noting that  $|(M \cap [0, k^{2\nu}])| \geq |(M_1 \cap [0, k^\nu])| \cdot |(M_{2,\nu})|$  gives then immediately the desired result.

Thus, it just remains to show that  $M_1$  and  $M_2$  have density 1.

Therefore we define  $M_{1,q'} := \{n \in \mathbb{N} \mid \delta'(q', (n)_k) \in \cup Q'_i\}$  and see that  $M_1 = \cap M_{1,q'}$ . Thus it suffices to show that  $M_{1,q'}$  has density 1.

One finds easily that for each  $q'_1 \in Q'$  there exists  $\mathbf{w}_{q'_1} \in \{0, \dots, k-1\}^*$  such that  $\delta'(q'_1, \mathbf{w}_{q'_1})$  belongs to  $Q'_i$  for some  $i$ . We take  $m$  to be the maximum of these lengths. As each  $Q'_i$  is closed under  $\delta'$  we may assume that each of these paths  $\mathbf{w}_{q'}$  is exactly of length  $m$ . We split  $Q'$  into two different parts, i.e.,  $\overline{Q'} = \cup Q_i, \widetilde{Q'} = Q' \setminus \overline{Q'}$ . We show that

$$(19) \quad \left| \{ \mathbf{w} \in \{0, \dots, k-1\}^n : \delta'(q', \mathbf{w}) \in \widetilde{Q'} \} \right| \leq k^m k^{n(1-\eta)}$$

where  $\eta$  is defined by  $\eta = 1 - \log(k^m - 1) / \log(k^m) > 0$  so that

$$k^m - 1 = k^{m(1-\eta)}.$$

We show (19) by induction on  $n$ . The statement is trivial for  $n \leq m$ . We find for  $n > m$  (as  $\overline{Q'}$  is closed under  $\delta'$ )

$$\begin{aligned} & |\{ \mathbf{w} \in \Sigma^n : \delta'(q', \mathbf{w}) \in \widetilde{Q'} \}| = \\ &= \sum_{q'_1 \in \widetilde{Q'}} |\{ \mathbf{w}_1 \in \Sigma^{n-m} : \delta'(q', \mathbf{w}) = q'_1 \}| \cdot |\{ \mathbf{w}_2 \in \Sigma^m : \delta'(q'_1, \mathbf{w}_2) \in \widetilde{Q'} \}| \\ &\leq \sum_{q'_1 \in \widetilde{Q'}} |\{ \mathbf{w}_1 \in \Sigma^{n-m} : \delta'(q', \mathbf{w}) = q'_1 \}| \cdot (k^m - 1) \\ &= |\{ \mathbf{w} \in \Sigma^{n-m} : \delta'(q', \mathbf{w}) \in \widetilde{Q'} \}| \cdot k^{m(1-\eta)} \leq k^m k^{(n-m)(1-\eta)} k^{m(1-\eta)}. \end{aligned}$$

[7, Lemma 2.2] shows

$$\lim_{\nu \rightarrow \infty} \frac{M_{2,\nu}}{k^\nu} = 1,$$

which finishes the proof. □

**5.2. Limiting distribution of  $T$  along primes.** The following proof shows that  $(T(q_0, (p)_k))_{p \in \mathcal{P}(a, k^\lambda)}$  has a limiting distribution.

Therefore, we need the following very important result from representation theory.

**Lemma 5.1.** *Let  $G$  be a compact group and  $\nu$  a regular normed Borel measure in  $G$ . Then a sequence  $(x_n)_{n \geq 0}$  is  $\nu$ -uniformly distributed in  $G$ , i.e.,  $\frac{1}{N} \sum_{n < N} \delta_{x_n} \rightarrow \nu$ , if and only if*

$$(20) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n < N} D(x_n) = \int_G D d\nu$$

holds for all irreducible unitary representations  $D$  of  $G$ .

A proof is given for example in [27].

**Lemma 3.7.** *Let  $\lambda \in \mathbb{N}$ ,  $\mathcal{T}$  be a naturally induced transducer with function  $T$ . For every  $g \in G$  there exists  $f_g$  such that*

$$\frac{1}{\pi(x; a, k^\lambda)} \sum_{\substack{p < x \\ p \equiv a \pmod{k^\lambda}}} \mathbf{1}_{[T(q_0, (p)_k) = g]} = f_g + o(1) \text{ for } x \rightarrow \infty$$

*holds (uniformly) for all  $a < k^\lambda$  such that  $(a, k) = 1$ .*

*Proof.* The sequence we are concerned with is  $(T(q_0, (p)_k))_{p \in \mathcal{P}(a, k^\lambda)}$ .

We start by computing the left side of (20) for  $D = D_\ell$  for our specific sequence. Recall that  $d' | k - 1$  holds.

$$\begin{aligned} & \lim_{x \rightarrow \infty} \frac{1}{\pi(x; a, k^\lambda)} \sum_{\substack{p < x \\ p \equiv a \pmod{k^\lambda}}} D_\ell(T(q_0, (p)_k)) \\ &= \lim_{x \rightarrow \infty} \frac{1}{\pi(x; a, k^\lambda)} \sum_{\substack{p < x \\ p \equiv a \pmod{k^\lambda}}} e(\ell p / d') \\ &= \lim_{x \rightarrow \infty} \frac{1}{\pi(x; a, k^\lambda)} \sum_{b < d'} e(\ell b / d') \sum_{\substack{p < x \\ p \equiv a \pmod{k^\lambda} \\ p \equiv b \pmod{d'}} 1 \\ &= \frac{1}{\varphi(d')} \sum_{\substack{b < d' \\ (b, d') = 1}} e(\ell b / d') \\ &= \frac{1}{\varphi(d')} S(\ell, 0; d'). \end{aligned}$$

Here  $S(a, b; c)$  denotes the Kloosterman (or Ramanujan) sum, defined by

$$S(a, b; c) := \sum_{\substack{x < c \\ (x, c) = 1}} e\left(\frac{ax + b\bar{x}}{c}\right),$$

where  $\bar{x}$  denotes the multiplicative inverse of  $x$  modulo  $c$ .

Furthermore, we want to show that

$$\lim_{x \rightarrow \infty} \frac{1}{\pi(x; a, k^\lambda)} \sum_{\substack{p < x \\ p \equiv a \pmod{k^\lambda}}} D(T(q_0, (p)_k)) = 0$$

holds for  $D \notin \{D_\ell : \ell = 0, \dots, d' - 1\}$ . We use a standard technique of analytic number theory to restrict ourself to sums with  $x = k^\nu$ . Let  $\nu$  be the unique integer such that

$k^{\nu-1} \leq x < k^\nu$ . We find (see for example [18])

$$\left\| \frac{1}{\pi(x; a, k^\lambda)} \sum_{\substack{p \leq x \\ p \equiv a \pmod{k^\lambda}}} D(T(q_0, (p)_k)) \right\|_F \ll \nu \sup_{\theta \in \mathbb{R}} \left\| \frac{1}{\pi(x; a, k^\lambda)} \sum_{\substack{p \leq k^\nu \\ p \equiv a \pmod{k^\lambda}}} e(\theta p) D(T(q_0, (p)_k)) \right\|_F.$$

We can rewrite the condition  $p \equiv a \pmod{k^\lambda}$  using exponential sums and find that the right side is bounded by

$$\begin{aligned} & \nu \frac{1}{\pi(k^{\nu-1}; a, k^\lambda)} \sup_{\theta \in \mathbb{R}} \left\| \frac{1}{k^\lambda} \sum_{h < k^\lambda} \sum_{p \leq k^\nu} e\left(h \frac{p-a}{k^\lambda}\right) e(\theta p) D(T(q_0, (p)_k)) \right\|_F \\ & \leq \nu \frac{1}{\pi(k^{\nu-1}; a, k^\lambda)} \sup_{\theta \in \mathbb{R}} \max_{h < k^\lambda} \left\| \sum_{p \leq k^\nu} e\left(p \left(\frac{h}{k^\lambda} + \theta\right)\right) D(T(q_0, (p)_k)) \right\|_F \\ & \leq \nu \frac{1}{\pi(k^{\nu-1}; a, k^\lambda)} \sup_{\theta' \in \mathbb{R}} \left\| \sum_{p \leq k^\nu} e(p\theta') D(T(q_0, (p)_k)) \right\|_F. \end{aligned}$$

Thus we find – by using Proposition 3.6 – that

$$\lim_{x \rightarrow \infty} \frac{1}{\pi(x; a, k^\lambda)} \sum_{\substack{p < x \\ p \equiv a \pmod{k^\lambda}}} D(T(q_0, (p)_k)) = \begin{cases} \frac{1}{\varphi(d')} S(\ell, 0; d'), & \text{for } D = D_\ell \\ 0, & \text{otherwise,} \end{cases}$$

We define the function  $f$  by

$$\begin{aligned} f(g) &:= 1 + \frac{1}{\varphi(d')} \overline{S(1, 0; d')} D_1(g) + \dots + \frac{1}{\varphi(d')} \overline{S(d'-1, 0; d')} D_{d'-1}(g) \\ &= \sum_{\ell=0}^{d'-1} \frac{1}{\varphi(d')} \overline{S(\ell, 0; d')} D_\ell(g). \end{aligned}$$

We recall that  $D_\ell(g) := e\left(\frac{so(g)\ell}{d'}\right)$  and, therefore, we find for  $v < d'$ ,  $D_\ell(T(nd' + v)) = e\left(\frac{v\ell}{d'}\right)$ .

By using this fact, we can determine  $f(T(nd', v))$ .

$$\begin{aligned} f(T(nd' + v)) &= \frac{1}{\varphi(d')} \sum_{\ell=0}^{d'-1} \overline{S(\ell, 0; d')} e\left(\frac{\ell v}{d'}\right) \\ &= \frac{1}{\varphi(d')} \sum_{\ell=0}^{d'-1} \sum_{\substack{x < d' \\ (x, d')=1}} e\left(-\frac{\ell x}{d'}\right) e\left(\frac{\ell v}{d'}\right) \\ &= \frac{1}{\varphi(d')} \sum_{\substack{x < d' \\ (x, d')=1}} \sum_{\ell=0}^{d'-1} e\left(\ell \frac{v-x}{d'}\right) \\ &= \frac{1}{\varphi(d')} \sum_{\substack{x < d' \\ (x, d')=1}} d' \mathbf{1}_{[x=v]}. \end{aligned}$$

In total we find

$$f(T(nd' + v)) = \begin{cases} \frac{d'}{\varphi(d')}, & \text{for } (v, d') = 1 \\ 0, & \text{otherwise,} \end{cases}$$

i.e.,  $f$  is a positive function. We can actually rewrite this by using Theorem 2.16:

$$f(g) = \begin{cases} \frac{d'}{\varphi(d')}, & \text{for } (s_0(g), d') = 1 \\ 0, & \text{otherwise,} \end{cases}.$$

Thus we can define

$$d\nu = f d\mu,$$

where  $\mu$  denotes the Haar-measure of  $G$ .

Our goal is to show that  $(T(q_0, (p)_k))_{p \in \mathcal{P}(a, k^\lambda)}$  is  $\nu$ -uniformly distributed in  $G$ .

Let  $\{D^\alpha = (d_{ij}^\alpha)_{i,j \leq n_\alpha}, \alpha \in \mathcal{A}\}$  be a complete set of pairwise inequivalent irreducible unitary representations and set  $e_{ij}^\alpha(g) = \sqrt{n_\alpha} d_{ij}^\alpha(g)$ . Note that  $A$  is finite as we are actually considering a finite group  $G$ . Recall that the set  $\{e_{ij}^\alpha\}$  forms a complete orthonormal system in the Hilbert space  $L^2(G)$ . We obtain for  $D_\ell$  that

$$\int_G D_\ell f d\mu = \frac{1}{\varphi(d')} \sum_{m < d'} S(m, 0; d') \langle D_\ell | D_m \rangle = \frac{1}{\varphi(d')} S(\ell, 0; d').$$

For all other representations  $D^\alpha = (d_{ij}^\alpha)_{1 \leq i, j \leq n_\alpha}$ , we find

$$\int_G d_{ij}^\alpha f d\mu = \frac{1}{\varphi(d')} \sum_{m < d'} S(m, 0; d') \langle d_{ij}^\alpha | D_m \rangle = 0.$$

This proves that  $(T(q_0, (p)_k))_{p \in \mathcal{P}(a, k^\lambda)}$  is  $\nu$ -uniformly distributed, where  $\nu$  does not depend on  $a$ .  $\square$

**5.3. Relation of Fourier terms.** We prove the following relation between very similarly defined terms that appeared when we estimated the Fourier terms.

**Lemma 4.6.** *Let  $D$  be a unitary representation of  $G$ . Assume that there exists  $\eta > 0$  such that*

$$\|\psi_{\lambda, \alpha}^q(t, r)\|_2 \ll k^{-\eta\lambda}$$

*holds uniformly for  $q \in Q$ ,  $\lambda, \alpha \in \mathbb{N}$ ,  $t \in \mathbb{R}$  and  $r < k^\alpha$ . Then Theorem 4.5 holds for  $D$ .*

*Proof.* We start by proving

$$(21) \quad \|\phi_{\lambda, \alpha}^q(t, r)\|_2 \leq \sum_{1 \leq j \leq \lambda} \max_{q' \in Q} \|\psi_{\lambda-j, \alpha}^{q'}(t, r \bmod k^\alpha)\|_2 \frac{1}{k^j} + \frac{1}{k^\lambda}.$$

We rewrite  $r = r_1 k^{\alpha+\lambda} + r_2 k^\alpha + r_3$  where  $r_3 < k^\alpha$  and  $r_2 < k^\lambda$ , such that  $uk^\alpha + r = r_1 k^{\alpha+\lambda} + (u + r_2)k^\alpha + r_3$ . We distinguish the cases  $r_1 > 0$  and  $r_1 = 0$  as this determines

whether the number of digits of  $(uk^\alpha + r)$  depends on  $u$ . Let us first consider the case  $r_1 > 0$ . We find by distinguishing  $u + r_2 < k^\lambda$  and  $u + r_2 \geq k^\lambda$

$$\begin{aligned}
 \phi_{\lambda,\alpha}^q(t, r) &= \frac{1}{k^\lambda} \sum_{\substack{u < k^\lambda \\ u+r_2 < k^\lambda}} D(T(q, (r_1)_k(u+r_2)_k^\lambda(r_3)_k^\alpha)) e(-ut) \\
 &\quad + \frac{1}{k^\lambda} \sum_{\substack{u < k^\lambda \\ u+r_2 \geq k^\lambda}} D(T(q, (r_1+1)_k(u+r_2-k^\lambda)_k^\lambda(r_3)_k^\alpha)) e(-ut) \\
 &= \frac{1}{k^\lambda} D(T(q, (r_1)_k)) \sum_{r_2 \leq u' < k^\lambda} D(T(\delta(q, (r_1)_k), (u')_k^\lambda(r_3)_k^\alpha)) e(-(u' - r_2)t) \\
 &\quad + \frac{1}{k^\lambda} D(T(q, (r_1+1)_k)) \sum_{u' < r_2} D(T(\delta(q, (r_1+1)_k), (u')_k^\lambda(r_3)_k^\alpha)) e(-(u' - r_2 + k^\lambda)t).
 \end{aligned}$$

This gives

$$\begin{aligned}
 \|\phi_{\lambda,\alpha}^q(t, r)\|_2 &\leq \max_{q' \in Q} \frac{1}{k^\lambda} \left\| \sum_{r_2 \leq u' < k^\lambda} D(T(q', (u')_k^\lambda(r_3)_k^\alpha)) e(-u't) \right\|_2 \\
 &\quad + \max_{q' \in Q} \frac{1}{k^\lambda} \left\| \sum_{u' < r_2} D(T(q', (u')_k^\lambda(r_3)_k^\alpha)) e(-u't) \right\|_2.
 \end{aligned}$$

The case  $r_1 = 0$  can be treated similarly and we find in this case

$$\begin{aligned}
 (22) \quad \|\phi_{\lambda,\alpha}^q(t, r)\|_2 &\leq \max_{q' \in Q} \frac{1}{k^\lambda} \left\| \sum_{r_2 \leq u' < k^\lambda} D(T(q', (u')_k(r_3)_k^\alpha)) e(-u't) \right\|_2 \\
 &\quad + \max_{q' \in Q} \frac{1}{k^\lambda} \left\| \sum_{u' < r_2} D(T(q', (u')_k(r_3)_k^\alpha)) e(-u't) \right\|_2.
 \end{aligned}$$

We work from now on just with the case  $r_1 = 0$  as the case  $r_1 > 0$  works similarly. We rewrite  $r_2 = r'_2 k^{\lambda-1} + r''_2$  with  $r'_2 < k$ ,  $r''_2 < k^{\lambda-1}$  and find by distinguishing the most significant digits of  $u'$  the following upper bound for the right side of (22):

$$\begin{aligned}
 &\frac{1}{k} \sum_{r'_2+1 \leq u'_1 < k} \max_{q' \in Q} \frac{1}{k^{\lambda-1}} \left\| \sum_{u'_2 < k^{\lambda-1}} D(T(q', (u'_2)_k^{\lambda-1}(r_3)_k^\alpha)) e(-u'_2 t) \right\|_2 \\
 &\quad + \frac{1}{k} \max_{q' \in Q} \frac{1}{k^{\lambda-1}} \left\| \sum_{r'_2 \leq u'_2 < k^{\lambda-1}} D(T(q', (u'_2)_k(r_3)_k^\alpha)) e(-u'_2 t) \right\|_2 \\
 &\quad + \frac{1}{k} \sum_{u'_1 < r'_2} \max_{q' \in Q} \frac{1}{k^{\lambda-1}} \left\| \sum_{u'_2 < k^{\lambda-1}} D(T(q', (u'_2)_k^{\lambda-1}(r_3)_k^\alpha)) e(-u'_2 t) \right\|_2
 \end{aligned}$$

$$+ \frac{1}{k} \max_{q' \in Q} \frac{1}{k^{\lambda-1}} \left\| \sum_{u'_2 < r'_2} D(T(q', (u'_2)_k^{\lambda-1} (r_3)_k^\alpha)) e(-u'_2 t) \right\|_2.$$

The first and third lines give

$$\frac{k-1}{k} \psi_{\lambda-1, \alpha}^q(t, r_3)$$

and we find in total that the right side of Eq. (22) is bounded by

$$\begin{aligned} & \frac{k-1}{k} \psi_{\lambda-1, \alpha}^q(t, r_3) + \frac{1}{k} \max_{q' \in Q} \frac{1}{k^{\lambda-1}} \left\| \sum_{r'_2 \leq u'_2 < k^{\lambda-1}} D(T(q', (u'_2)_k (r_3)_k^\alpha)) e(-u'_2 t) \right\|_2 \\ & + \frac{1}{k} \max_{q' \in Q} \frac{1}{k^{\lambda-1}} \left\| \sum_{u'_2 < r'_2} D(T(q', (u'_2)_k^{\lambda-1} (r_3)_k^\alpha)) e(-u'_2 t) \right\|_2. \end{aligned}$$

By applying this step inductively we find for  $1 \leq \ell \leq \lambda$  the bound

$$\begin{aligned} & \sum_{1 \leq j \leq \ell} \frac{k-1}{k^j} \psi_{\lambda-j, \alpha}^q(t, r_3) + \frac{1}{k^\ell} \max_{q' \in Q} \frac{1}{k^{\lambda-\ell}} \left\| \sum_{r_2^{(\ell)} \leq u'_2 < k^{\lambda-\ell}} D(T(q', (u'_2)_k (r_3)_k^\alpha)) e(-u'_2 t) \right\|_2 \\ & + \frac{1}{k^\ell} \max_{q' \in Q} \frac{1}{k^{\lambda-\ell}} \left\| \sum_{u'_2 < r_2^{(\ell)}} D(T(q', (u'_2)_k^{\lambda-\ell} (r_3)_k^\alpha)) e(-u'_2 t) \right\|_2. \end{aligned}$$

(21) is obtained by taking  $\ell = \lambda$ . Thus it just remains to use the bound of  $\psi_{\lambda, \alpha}^q(t, r)$

$$\begin{aligned} \|\phi_{\lambda, \alpha}^q(t, r)\|_2 & \leq \sum_{1 \leq j \leq \lambda} \max_{q' \in Q} \left\| \psi_{\lambda-j, \alpha}^{q'}(t, r \bmod k^\alpha) \right\|_2 \frac{1}{k^j} + \frac{1}{k^\lambda} \\ & \leq c \sum_{j \leq \lambda} k^{-(\lambda-j)\eta} k^{-j} + k^{-\lambda} \\ & \leq ck^{-\lambda\eta} \frac{k^{-(1-\eta)}}{1 - k^{-(1-\eta)}} + k^{-\lambda}. \end{aligned}$$

□

**5.4. Proof of Theorem 4.4 and Theorem 4.3.** The proof of Theorem 4.3 and Theorem 4.4 is completely analogous to the corresponding proof in [31] and relies on bounds for sums of type I and sums of type II. In this part of the paper we will comment on how to adapt the original proof to our situation. We will only describe the important changes briefly (and assume that the reader is familiar with [31]). First we comment on how to adapt the auxiliary results of [31].

For a generalization of [31, Lemma 3] see [10]. [31, Lemma 6] as well as the Cauchy-Schwarz inequality can be easily adapted to matrix-valued functions, where we use the Frobenius norm instead of the absolute value. These results change at most by a factor  $\sqrt{d}$ .

[31, Lemma 8] can be adapted to our case where the proof stays unchanged:



**Lemma 5.2.** *Let  $f : \mathbb{N} \rightarrow U_d$  satisfy Definition 4.1 with  $\eta > 0$ . For  $(\mu, \nu, \rho) \in \mathbb{N}^3$  with  $2\rho < \nu$  the set  $\mathcal{E}$  of pairs  $(m, n) \in \{k^{\mu-1}, \dots, k^\mu - 1\} \times \{k^{\nu-1}, \dots, k^\nu - 1\}$  such that there exists  $\ell < k^{\mu+\rho}$  with  $f(mn+k)^H f(mn) \neq f_{\mu+2\rho}(mn+k)^H f_{\mu+2\rho}(mn)$  satisfies*

$$(23) \quad \text{card } \mathcal{E} \ll (\log k) k^{\mu+\nu-\eta\rho}.$$

For  $a \in \mathbb{Z}$  and  $\kappa \in \mathbb{N}$  we denote by  $\mathbf{r}_\kappa(a)$  the unique integer  $r \in \{0, \dots, k^\kappa - 1\}$  such that  $a \equiv r \pmod{k^\kappa}$ . More generally for integers  $0 \leq \kappa_1 \leq \kappa_2$  we denote by  $\mathbf{r}_{\kappa_1, \kappa_2}(a)$  the unique integer  $r \in \{0, \dots, k^{\kappa_2 - \kappa_1} - 1\}$  such that  $a = sk^{\kappa_2} + rk^{\kappa_1} + t$  for some  $t \in \{0, \dots, k^{\kappa_1} - 1\}$  and  $s \in \mathbb{Z}$ . [31, Lemma 9] can be adapted as well:

**Lemma 5.3.** *Let  $f : \mathbb{N} \rightarrow U_d$  satisfying Definition 4.1 and  $(\mu, \nu, \mu_0, \mu_1, \mu_2) \in \mathbb{N}^5$  with  $\mu_0 \leq \mu_1 \leq \mu \leq \mu_2$ ,  $\mu \leq \nu$  and  $2(\mu_2 - \mu) \leq \mu_0$ . For  $(a, b, c) \in \mathbb{N}^3$  the set  $\mathcal{E}(a, b, c)$  of pairs  $(m, n) \in \{k^{\mu-1}, \dots, k^\mu - 1\} \times \{k^{\nu-1}, \dots, k^\nu - 1\}$  such that*

$$\begin{aligned} & f_{\mu_2}(mn + am + bn + c)^H f_{\mu_2}(k^{\mu_0} \mathbf{r}_{\mu_0, \mu_2}(mn + am + bn + c)) \\ & \neq f_{\mu_1}(mn + am + bn + c)^H f_{\mu_1}(k^{\mu_0} \mathbf{r}_{\mu_0, \mu_2}(mn + am + bn + c)) \end{aligned}$$

satisfies

$$(24) \quad \text{card } \mathcal{E}(a, b, c) \ll \max(\tau(k), \log k) \mu_2^{\omega(k)} k^{\mu+\nu+\eta(\mu_0-\mu_1)}.$$

5.4.1. *Sums of Type I.* We take a non-decreasing function  $\gamma : \mathbb{R} \rightarrow \mathbb{R}$  satisfying  $\lim_{\lambda \rightarrow \infty} \gamma(\lambda) = +\infty$ ,  $c \geq 2$  and  $f : \mathbb{N} \rightarrow U_d$  be a function satisfying Definition 4.1 and  $f \in \mathcal{F}_{\gamma, c}$  according to Definition 4.2. Let

$$(25) \quad 1 \leq M \leq N \text{ such that } M \leq (MN)^{1/3}.$$

Let  $\mu$  and  $\nu$  be the unique integers such that

$$k^{\mu-1} \leq M < k^\mu \text{ and } k^{\nu-1} \leq N < k^\nu.$$

Let  $\theta \in \mathbb{R}$ , an interval  $I(M, N) \subseteq [0, MN]$  and

$$(26) \quad S_I(\theta) = \sum_{\frac{M}{q} < m \leq M} \left\| \sum_{n: mn \in I(M, N)} f(mn) e(\theta mn) \right\|_F.$$

**Proposition 5.4.** *Assuming (25) and with  $c \geq 2$ , we have – uniformly for  $\theta \in \mathbb{R}$  – that*

$$(27) \quad S_I(\theta) \ll (\log k)^{5/2} (\mu + \nu)^2 k^{\mu+\nu-\frac{\eta}{2}\gamma(\frac{\mu+\nu}{3})}.$$

*Proof (Sketch):* The proof proceeds as in [31] up to [31, equation (32)]. We find

$$\begin{aligned} \widehat{f_{\mu+\nu}}(t) &= \frac{1}{k^{\mu+\nu-\alpha}} \sum_{v < k^{\mu+\nu-\alpha}} f(vk^\alpha) e\left(-\frac{vt}{k^{\mu+\nu-\alpha}}\right) \\ &\quad \frac{1}{k^\alpha} \sum_{u \leq k^\alpha} f(vk^\alpha)^H f(u + vk^\alpha) e\left(-\frac{ut}{k^{\mu+\nu}}\right). \end{aligned}$$

Here we needed to ensure the correct order of the terms. From now on the proof does not change and we just have to take the factor  $\eta$  into account when using Definition 4.1. We find for example

$$\text{card } \widetilde{\mathcal{W}}_\alpha \ll k^{\mu+\nu-\eta\rho_1}.$$

Thereafter one only needs to keep the order of terms as the values of  $f$  are now matrices and thus the order of terms is important. However, this does not change any important properties and all arguments of [31] still hold. By taking  $\eta$  into account, we find

$$\begin{aligned} \|S''_{I,2}(M, d)\|_F &\ll (\log k)^{1/2} k^{\mu+\nu-\eta\rho_1/2} \\ S'_{I,2}(\theta') &\ll (\log k)^{1/2} \sum_{1 \leq d \leq M} \frac{k^{-\eta\rho_1/2}}{d} \ll \mu(\log k)^{3/2} k^{-\eta\rho_1/2}. \end{aligned}$$

Choosing

$$\rho_1 = \gamma \left( \frac{\mu + \nu}{3} \right) \frac{2}{1 + \eta}$$

then gives the desired result.  $\square$

**5.4.2. Sums of Type II.** We take  $\gamma : \mathbb{R} \rightarrow \mathbb{R}$  a non-decreasing function satisfying  $\lim_{\lambda \rightarrow \infty} \gamma(\lambda) = +\infty$ ,  $c \geq 10$  and  $f : \mathbb{N} \rightarrow U_d$  be a function satisfying Definition 4.1 and  $f \in \mathcal{F}_{\gamma, c}$  in Definition 4.2. Let  $M, N$  such that  $1 \leq M \leq N$ . We let  $\mu$  and  $\nu$  denote the unique integers such that

$$k^{\mu-1} \leq M < k^\mu \text{ and } k^{\nu-1} \leq N < k^\nu.$$

Let us assume that

$$(28) \quad \frac{1}{4}(\mu + \nu) \leq \mu \leq \nu \leq \frac{3}{4}(\mu + \nu).$$

We assume – as in [31] – that the multiplicative dependency of the variables in the type II sums has been removed by the classical method described (for example) in [30, Section 5]. Let  $\theta \in \mathbb{R}$ ,  $a_m \in \mathbb{C}$ ,  $b_n \in \mathbb{C}$  with  $|a_m| \leq 1$ ,  $|b_n| \leq 1$  and

$$S_{II}(\theta) = \sum_m \sum_n a_m b_n f(mn) e(\theta mn)$$

where we sum over  $m \in (M/k, M]$  and  $n \in (N/k, N]$ . We will prove

**Proposition 5.5.** *Assuming (28) and  $c \geq 10$ , uniformly for  $|a_m| \leq 1$ ,  $|b_n| \leq 1$  and  $\theta \in \mathbb{R}$ , we have*

$$(29) \quad \|S_{II}(\theta)\|_F \ll \max(\tau(k) \log k, \log^3 k)^{1/4} (\mu + \nu)^{\frac{1}{4}(1 + \max(\omega(k), 2))} k^{\mu+\nu - \frac{\gamma}{20} \gamma(2 \lfloor \mu/15 \rfloor)}.$$

*Proof.* The proof of the corresponding result in [31] is the most difficult part – quite long and complicated. We will try to focus only on the necessary changes and keep it as short as possible, as no important new ideas are needed.

We find by using the corresponding results:

$$\|S_{II}(\theta)\|_F^2 \ll \frac{M^2 N^2}{R} + \frac{MN}{R} \sum_{1 \leq r < R} \left(1 - \frac{r}{R}\right) \text{tr}(S_1(r))$$

with

$$S_1(r) = \sum_m \sum_{n \in I(N, r)} b_{n+r} \overline{b_n} f(mn + mr) f(mn)^H e(\theta mr),$$

where  $I(N, r) = (N/k, N - r]$ . Let

$$(30) \quad \mu_2 = \mu + 2\rho.$$

As we are only interested in  $\text{tr}(S_1(r))$ , we can – by well-known properties of the trace – exchange the order of the matrices

$$(31) \quad \text{tr}(S_1(r)) = \sum_m \sum_{n \in I(N,r)} b_{n+r} \bar{b}_n \text{tr}(f(mn + mr) f(mn)^H) e(\theta mr)$$

$$(32) \quad = \sum_m \sum_{n \in I(N,r)} b_{n+r} \bar{b}_n \text{tr}(f(mn)^H f(mn + mr)) e(\theta mr)$$

$$(33) \quad = \text{tr} \left( \sum_m \sum_{n \in I(N,r)} b_{n+r} \bar{b}_n f(mn)^H f(mn + mr) e(\theta mr) \right).$$

We denote from now on by  $\tilde{S}$  the corresponding sum where we exchange the order of the matrices, e.g. :

$$(34) \quad \tilde{S}_1(r) = \sum_m \sum_{n \in I(N,r)} b_{n+r} \bar{b}_n f(mn)^H f(mn + mr) e(\theta mr).$$

If  $f$  satisfies the carry property described in Definition 4.1, then by Lemma 5.2 the number of pairs  $(m, n)$  for which  $f(mn)^H f(mn + mr) \neq f_{\mu_2}(mn)^H f_{\mu_2}(mn + mr)$  is  $O(k^{\mu+\nu-\eta\rho})$ . Hence

$$(35) \quad \text{tr}(S_1(r)) = \text{tr}(\tilde{S}_1(r)) = \text{tr}(S'_1(r)) + O(k^{\mu+\nu-\eta\rho})$$

where

$$S'_1(r) = \sum_m \sum_{n \in I(N,r)} b_{n+r} \bar{b}_n f_{\mu_2}(mn)^H f_{\mu_2}(mn + mr) e(\theta mr).$$

Note that it was important to change the order of matrices first to apply Lemma 5.2. Using again the Cauchy-Schwarz inequality for the summation over  $r$  and  $|\text{tr}(A)| \leq \|A\|_F \sqrt{d}$ , leads to

$$(36) \quad \|S_{II}(\theta)\|_F^4 \ll \frac{M^4 N^4}{R^2} + \frac{M^2 N^2}{R^2} R \sum_{1 \leq r < R} \|S'_1(r)\|_F^2.$$

When applying the Van-der-Corput inequality for the summation over  $m$  (see for example [10]) we need again to keep the correct order of terms. We find

$$\sum_{1 \leq r < R} \|S'_1(r)\|_F^2 \ll \frac{M^2 N^2 R}{S} + \frac{MN}{S} \text{tr}(\tilde{S}_2)$$

with

$$\tilde{S}_2 = \sum_{1 \leq r < R} \sum_{1 \leq s < S} \left(1 - \frac{s}{S}\right) e(\theta k^{\mu_1} r s) \tilde{S}'_2(r, s)$$

where

$$\begin{aligned} \tilde{S}'_2(r, s) &= \sum_m \sum_n f_{\mu_2}((m + sk^{\mu_1})n)^H f_{\mu_2}((m + sk^{\mu_1})(n + r)) f_{\mu_2}(m(n + r))^H f_{\mu_2}(mn) \\ &= \sum_m \sum_n f_{\mu_1}((m + sk^{\mu_1})n)^H f_{\mu_1, \mu_2}((m + sk^{\mu_1})n)^H f_{\mu_1, \mu_2}((m + sk^{\mu_1})(n + r)) \\ &\quad f_{\mu_1}((m + sk^{\mu_1})(n + r)) f_{\mu_1}(m(n + r))^H f_{\mu_1, \mu_2}(m(n + r))^H f_{\mu_1, \mu_2}(mn) f_{\mu_1}(mn). \end{aligned}$$

However, we find that  $\text{tr}(\tilde{S}'_2(r, s)) = \text{tr}(S'_2(r, s))$  for

$$S'_2(r, s) = \sum_m \sum_n f_{\mu_1, \mu_2}(mn) f_{\mu_1, \mu_2}((m + sk^{\mu_1})n)^H f_{\mu_1, \mu_2}((m + sk^{\mu_1})(n + r)) f_{\mu_1, \mu_2}(m(n + r))^H$$

and, therefore,

$$\sum_{1 \leq r < R} \|S'_1(r)\|_F^2 \ll \frac{M^2 N^2 R}{S} + \frac{MN}{S} \text{tr}(S_2)$$

where

$$S_2 = \sum_{1 \leq r < R} \sum_{1 \leq s < S} \left(1 - \frac{s}{S}\right) e(\theta k^{\mu_1} r s) S'_2(r, s).$$

We choose the order of the terms so that we can use the Cauchy-Schwarz inequality efficiently. Whenever we use Definition 4.1 (at least implicitly), we find a different error term, e.g., instead of [31, (60)] we can use

$$\text{card } \mathcal{E}_{\mu_0, \mu_1, \mu_2}(r, s) \ll \max(\tau(k), \log k) (\mu + \nu)^{\omega(k)} k^{\mu + \nu - 2\eta\rho'}.$$

Thus we find

$$S'_2(r, s) = S_3(r, s) + O(\max(\tau(k), \log k) (\mu + \nu)^{\omega(k)} k^{\mu + \nu - 2\eta\rho'}),$$

where the order of the terms in  $S_3$  has to be changed:

$$S_3(r, s) = \sum_m \sum_n \sum_{0 \leq u_0 < k^{\mu_2 - \mu_0}} \sum_{0 \leq u_1 < k^{\mu_2 - \mu_0}} \chi_{k^{\mu_0 - \mu_2}} \left( \frac{mn}{k^{\mu_2}} - \frac{u_0}{k^{\mu_2 - \mu_0}} \right) \chi_{k^{\mu_0 - \mu_2}} \left( \frac{mn + mr}{k^{\mu_2}} - \frac{u_1}{k^{\mu_2 - \mu_0}} \right) \\ g(u_0) g(u_0 + k^{\mu_1 - \mu_0} sn)^H g(u_1 + k^{\mu_1 - \mu_0} sn + k^{\mu_1 - \mu_0} sr) g(u_1)^H.$$

This impact of  $\eta$  carries through the rest of the work and we will only comment on the specific form of some intermediate results, e.g.,  $S_4$ :

$$S_4(r, s) = k^{2(\mu_2 - \mu_0)} \sum_{|h_0| \leq H} \sum_{|h_1| \leq H} a_{h_0}(k^{\mu_0 - \mu_2}, H) a_{h_1}(k^{\mu_0 - \mu_2}, H) \sum_{0 \leq h_2 < k^{\mu_2 - \mu_0}} \sum_{0 \leq h_3 < k^{\mu_2 - \mu_0}} e \left( \frac{h_3 sr}{k^{\mu_2 - \mu_1}} \right) \\ \widehat{g}(h_0 - h_2) \widehat{g}(-h_2)^H \widehat{g}(h_3) \widehat{g}(h_3 - h_1)^H \\ \sum_m \sum_n e \left( \frac{(h_0 + h_1)mn + h_1 mr + (h_2 + h_3)k^{\mu_1} sn}{k^{\mu_2}} \right).$$

The definitions of  $S_6$  and  $S_7$  have to be adapted as well, e.g. :

$$S_7(h_1) = \sum_{0 \leq h' < k^{\mu_2 - \mu_0}} \|\widehat{g}(h' - h_1) \widehat{g}(h')^H\|_F^2.$$

The following arguments of [31] carry over to this generalization.

We find the following lemma, which is the analogue to [31, Lemma 10].

**Lemma 5.6.** *If*

$$\mu \leq \left(2 + \frac{4}{3}c\right) \rho$$

then we have – uniformly for  $\lambda \in \mathbb{N}$  with  $\frac{1}{3}(\mu_2 - \mu_0) \leq \lambda \leq \frac{4}{5}(\mu_2 - \mu_0)$  –

$$\sum_{0 \leq h < k^{\mu_2 - \mu_0}} \sum_{0 \leq k < k^{\mu_2 - \mu_0 - \lambda}} \|\widehat{g}(h+k)\widehat{g}(h)^H\|_F^2 \ll k^{-\gamma_1(\lambda, \mu_1 - \mu_0)} (\log k^{\mu_2 - \mu_1})^2$$

where

$$(37) \quad \gamma_1(\lambda, \mu_1 - \mu_0) = \frac{\gamma(\lambda) - \mu_1 + \mu_0}{2} \eta.$$

*Proof.* The proof works as in [31], where one has to be careful to keep the order of terms. However, this does not change the proof substantially and by using the new estimate given by Definition 4.1 one finds the desired result.  $\square$

The rest of the proof does not change and it just remains to balance the error terms differently. One finds in total, uniformly for  $\theta \in \mathbb{R}$

$$\begin{aligned} \|S_{II}(\theta)\|_F^4 &\ll k^{4\mu+4\nu+\mu_1-\mu_0} (k^{-\gamma_1(\mu_2-\mu_0-2\rho, \mu_1-\mu_0)} + k^{-\rho} \log k^\rho) \\ &\quad (\tau(k^{\mu_2-\mu_1}) + k^{\mu_2-\mu_1-\nu \log k^{\mu_2-\mu_1}}) \\ &\quad + (\log k)^3 (\mu + \nu)^3 k^{4\mu+4\nu+3(\mu_2-\mu_0)+2\rho} (k^{-\mu_2} + k^{-\nu}) \\ &\quad + \max(\log k^{\mu_0}, \tau(k^{\mu_0})) k^{4\mu+4\nu-2\rho} \\ &\quad + \max(\tau(k), \log k) (\mu + \nu)^{\omega(k)} k^{4\mu+4\nu-2\eta\rho'}. \end{aligned}$$

Note that only the first and last error terms have changed compared to [31]. As in [31], we assume

$$\begin{aligned} \mu_2 &= \mu + 2\rho \\ \mu_1 &= \mu - 2\rho \\ \mu_0 &= \mu_1 - 2\rho' \\ \mu &\leq \nu \leq 3\mu. \end{aligned}$$

In total, we find

$$\begin{aligned} \|S_{II}(\theta)\|_F^4 &\ll \tau(k) (\mu_2 - \mu_1)^{\omega(k)} k^{4\mu+4\nu+\frac{2+\eta}{2}(\mu_1-\mu_0)-\frac{\eta}{2}\gamma(2\rho)} \log k^\rho \\ &\quad + (\log k)^3 (\mu + \nu)^3 k^{4\mu+4\nu+3(\mu_1-\mu_0)+14\rho-\mu} \\ &\quad + \max(\log k^{\mu_0}, \tau(k^{\mu_0})) k^{4\mu+4\nu-2\rho} \\ &\quad + \max(\tau(k), \log k) (\mu + \nu)^{\omega(k)} k^{4\mu+4\nu-2\eta\rho'}. \end{aligned}$$

Taking

$$\rho' = \lfloor \eta\gamma(2\rho)/10 \rfloor,$$

we have  $\mu_1 - \mu_0 = 2\rho' \leq \eta\gamma(2\rho)/5 \leq \eta\rho/5$ , and thus

$$\frac{2+\eta}{2}(\mu_1 - \mu_0) - \frac{\eta}{2}\gamma(2\rho) \leq \frac{3}{10}\eta\gamma(2\rho) - \frac{\eta}{2}\gamma(2\rho) = -\eta\frac{\gamma(2\rho)}{5}.$$

Furthermore, we choose

$$\rho = \lfloor \mu/15 \rfloor,$$

which yields by the same arguments as in [31] that – finally –

$$\|S_{II}(\theta)\|_F^4 \ll \max(\tau(k) \log k, \log^3 k)(\mu + \nu)^{1+\max(\omega(k), 2)} k^{4\mu+4\nu-\eta\gamma(2\lfloor\mu/15\rfloor)/5},$$

which completes the proof of Proposition 5.5.  $\square$

*Proof.* Proof of Theorem 1.2 and Theorem 1.3: The proof is completely analogous to the corresponding proof in [31].  $\square$

## REFERENCES

- [1] J.-P. Allouche and J. Shallit. *Automatic Sequences. Theory, Applications, Generalizations*. Cambridge: Cambridge University Press, 2003.
- [2] J. Bourgain. Möbius-Walsh correlation bounds and an estimate of Mauduit and Rivat. *J. Anal. Math.*, 119:147–163, 2013.
- [3] J. Bourgain. On the correlation of the Moebius function with rank-one systems. *J. Anal. Math.*, 120:105–130, 2013.
- [4] J. Bourgain, P. Sarnak, and T. Ziegler. Disjointness of Moebius from horocycle flows. In *From Fourier analysis and number theory to Radon transforms and geometry*, volume 28 of *Dev. Math.*, pages 67–83. Springer, New York, 2013.
- [5] C. Dartyge and G. Tenenbaum. Sommes des chiffres de multiples d’entiers. *Ann. Inst. Fourier (Grenoble)*, 55(7):2423–2474, 2005.
- [6] H. Davenport. On some infinite series involving arithmetical functions (II). *The Quarterly Journal of Mathematics*, os-8(1):313–320, 1937.
- [7] J.-M. Deshouillers, M. Drmota, and C. Müllner. Automatic Sequences generated by synchronizing automata fulfill the Sarnak conjecture. *Studia Mathematica*, 231:83–95, 2015.
- [8] T. Downarowicz and S. Kasjan. Odometers and Toeplitz subshifts revisited in the context of Sarnak’s conjecture. *ArXiv e-prints*, Feb. 2015. Available at <https://arxiv.org/pdf/1502.02307.pdf>.
- [9] M. Drmota. Subsequences of automatic sequences and uniform distribution. In *Uniform Distribution and Quasi-Monte Carlo methods*, volume 15 of *Radon Series on Computational and Applied Mathematics*, pages 87–104. De Gruyter, Berlin, 2014.
- [10] M. Drmota and J. F. Morgenbesser. Generalized Thue-Morse sequences of squares. *Israel J. Math.*, 190:157–193, 2012.
- [11] E. H. El Abdalaoui, S. Kasjan, and M. Lemańczyk. 0-1 sequences of the Thue-Morse type and Sarnak’s conjecture. *Proc. Amer. Math. Soc.*, 144(1):161–176, 2016.
- [12] E. H. El Abdalaoui, M. Lemańczyk, and T. de la Rue. On spectral disjointness of powers for rank-one transformations and Möbius orthogonality. *J. Funct. Anal.*, 266(1):284–317, 2014.
- [13] S. Ferenczi. Complexity of sequences and dynamical systems. *Discrete Math.*, 206(1-3):145–154, 1999. *Combinatorics and number theory (Tiruchirappalli, 1996)*.
- [14] S. Ferenczi, J. Kułaga-Przymus, M. Lemanczyk, and C. Mauduit. Substitutions and Möbius disjointness. In *Ergodic Theory, Dynamical Systems, and the Continuing Influence of John C. Oxtoby*, Contemporary Mathematics. American Mathematical Society, 2016.
- [15] S. Ferenczi and C. Mauduit. On Sarnak’s conjecture and Veech’s question for interval exchanges. <http://iml.univ-mrs.fr/ferenczi/fm2.pdf>, Preprint, 2015.
- [16] N. P. Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.
- [17] N. Frantzikinakis. Ergodicity of the Liouville system implies the Chowla conjecture. *ArXiv e-prints*, Nov. 2016. Available at <https://arxiv.org/pdf/1611.09338.pdf>.
- [18] S. W. Graham and G. Kolesnik. *van der Corput’s method of exponential sums*, volume 126 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1991.
- [19] B. Green. On (not) computing the Möbius function using bounded depth circuits. *Combin. Probab. Comput.*, 21(6):942–951, 2012.

- [20] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math. (2)*, 175(2):541–566, 2012.
- [21] G. Hanna. Sur les occurrences des mots dans les nombres premiers. *Acta Arith.*, 178(1):15–42, 2017.
- [22] E. Houcein El Abdalaoui, J. Kulaga-Przymus, M. Lemanczyk, and T. De La Rue. The Chowla and the Sarnak conjectures from ergodic theory point of view. *ArXiv e-prints*, Oct. 2014. Available at <https://arxiv.org/pdf/1410.1673.pdf>.
- [23] E. Houcein El Abdalaoui, M. Lemanczyk, and T. De La Rue. Automorphisms with quasi-discrete spectrum, multiplicative functions and average orthogonality along short intervals. *ArXiv e-prints*, July 2015. Available at <https://arxiv.org/pdf/1507.04132.pdf>.
- [24] K.-H. Indlekofer and I. Kátai. Investigations in the theory of  $q$ -additive and  $q$ -multiplicative functions. I. *Acta Math. Hungar.*, 91(1-2):53–78, 2001.
- [25] D. Karagulyan. On Möbius orthogonality for interval maps of zero entropy and orientation-preserving circle homeomorphisms. *Ark. Mat.*, 53(2):317–327, 2015.
- [26] I. Kátai. A remark on a theorem of H. Daboussi. *Acta Math. Hungar.*, 47(1-2):223–225, 1986.
- [27] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974. Pure and Applied Mathematics.
- [28] J. Kułaga Przymus and M. Lemańczyk. The Möbius function and continuous extensions of rotations. *Monatsh. Math.*, 178(4):553–582, 2015.
- [29] J. Liu and P. Sarnak. The Möbius function and distal flows. *Duke Math. J.*, 164(7):1353–1399, 2015.
- [30] C. Mauduit and J. Rivat. Sur un problème de Gelfond : la somme des chiffres des nombres premiers. *Ann. of Math. (2)*, 171(3):1591–1646, 2010.
- [31] C. Mauduit and J. Rivat. Prime numbers along Rudin-Shapiro sequences. *J. Eur. Math. Soc. (JEMS)*, 17(10):2595–2642, 2015.
- [32] R. Peckner. Möbius disjointness for homogeneous dynamics. *ArXiv e-prints*, June 2015. Available at <https://arxiv.org/pdf/1506.07778.pdf>.
- [33] P. Sarnak. Three lectures on the Mobius function randomness and dynamics. <https://www.math.ias.edu/files/wam/2011/PSMobius.pdf>, 2011.
- [34] P. Sarnak and A. Ubis. The horocycle flow at prime times. *J. Math. Pures Appl. (9)*, 103(2):575–618, 2015.
- [35] J.-P. Serre. *Linear Representations of Finite Groups*. Springer-Verlag, New York-Heidelberg, 1977. Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42.
- [36] T. Tao. Möbius randomness of the Rudin-Shapiro sequence. <http://mathoverflow.net/questions/97261>, 2012.
- [37] T. Tao. Equivalence of the logarithmically averaged Chowla and Sarnak conjectures. *ArXiv e-prints*, May 2016. Available at <https://arxiv.org/pdf/1605.04628.pdf>.
- [38] W. A. Veech. Möbius orthogonality for generalized Morse-Kakutani flows. *American Journal of Mathematics*, 2016. (to appear).