Finite-State Transducers in the Wheeler Setting

Giovanna D'Agostino[1], Andrea Paradiso[1]

[1]University of Udine, DMIF, Udine, Italy

Abstract

We investigate the interaction between Wheeler automata and finite-state subsequential transducers. Wheeler languages form a structurally well-behaved subclass of the regular languages and provide an automata-theoretic foundation for efficient indexing and compression techniques inspired by the Burrows–Wheeler transform. We introduce a class of Wheeler transducers whose underlying input automata are Wheeler and whose output functions satisfy a piecewise monotonicity condition on the sets of words reaching each state. We show that this class is closed under composition and that inverse images of Wheeler languages under Wheeler transductions are again Wheeler languages. We then study minimization by refining Choffrut's syntactic congruence for subsequential functions with convexity and piecewise monotonicity constraints. This yields a Myhill–Nerode-style characterization of the functions realizable by Wheeler transducers. Finally, we provide a machine-independent characterization of Wheeler functions in terms of the behavior of monotone sequences and piecewise monotonicity. These results constitute a first step toward a structural theory of transductions compatible with the Wheeler setting.

Keywords

Wheeler automata, finite-state transducers, Wheeler languages

CEUR-WS.org/Vol-4269/paper46.pdf

1. Introduction

In this work we take a preliminary step toward understanding the connections between Wheeler automata and finite-state transducers.

Wheeler graphs form a class of edge-labeled directed graphs that admit a sorting of their nodes that is consistent with the co-lexicographic order of the labels of paths ending at them. A finite-state automaton whose underlying graph is a Wheeler graph is called a Wheeler automaton, and the languages recognized by such automata are called Wheeler languages [1]. Wheeler languages form a proper subclass of the regular languages; in fact, they are a subclass of the star-free languages. They have attracted considerable interest thanks to structural properties enabling highly efficient indexing, pattern matching and compression [2, 3, 4]. They provide a formal automata-theoretic foundation for data structures inspired by the Burrows–Wheeler transform, bridging language theory and practical text indexing. For this reason, Wheeler languages are important both as a mathematically natural subclass of regular languages and as a model for highly efficient algorithms on strings, graphs, and automata.

A natural way to deepen our understanding of a language class is to study the transformations that preserve it, either under direct image or under inverse image. This perspective has proved fruitful in the regular setting, where rational functions definable by finite-state subsequential transducers (FSTs, [5]) are known to preserve regularity under both images and inverse images, and where algebraic tools, such as syntactic invariants associated with transductions, provide structural information about the transformations involved. Subsequential transducers are a fundamental model for deterministic input-to-output transformations, capturing computations in which the output is produced online while the input is read from left to right. They are important both theoretically, as a well-behaved subclass of finite-state transductions, and practically, as efficient models for tasks such as text processing, coding, normalization, and stream transformation (see [6] for a recent survey).

Identifying classes of transducers that are compatible with the Wheeler structure may therefore help clarify the boundaries of Wheeler languages and provide a more dynamic, transformation-oriented

ICTCS 2026: Italian Conference on Theoretical Computer Science, September 7–9, 2026, Udine, Italy
giovanna.dagostino@uniud.it (G. D'Agostino); andrea.paradiso.uni@gmail.com (A. Paradiso)
0000-0002-8920-483X (G. D'Agostino); 0000-0002-3614-2487 (A. Paradiso)
© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-09-19

view of this class. In this work we define a subclass of transducers and argue that it is a natural class of transducers to consider in the setting of Wheeler languages. What properties should such a class have? We would certainly want it to be closed under composition, and, with respect to Wheeler languages, we would want it to be at least Wheeler-continuous, in the sense that the preimage of a Wheeler language for a transducer in the class should remain a Wheeler language [1]. In this work we indeed prove that our class is closed under composition and computes Wheeler-continuous functions. Moreover we give a machine-independent characterization of the functions computed by the class. Due to space limitations, all proofs are omitted. The interested reader can find them in the complete arXiv version entitled "Finite-State Transducers in the Wheeler Setting".

2. Preliminaries

We first fix some conventions before introducing Wheeler transducers in detail. Throughout the paper, whenever an alphabet Σ is given, it is assumed to be equipped with an order denoted by $\prec$. Moreover, given two strings $\alpha = a_1a_2 \dots a_n$ and $\beta = b_1b_2 \dots b_m$ in Σ^*, their co-lexicographic order is defined by overloading the symbol $\prec$:

$$\alpha \prec \beta \iff \big(n < m \wedge (\forall j < n)(a_{n-j} = b_{m-j})\big)\vee \\ (\exists i)(a_{n-i} \prec b_{m-i} \wedge (\forall j < i)(a_{n-j} = b_{m-j}))$$

(in other words, the co-lexicographic order coincides with the lexicographic one on the reverse of the words).
We assume that the notions of deterministic (DFA)/nondeterministic (NFA) automata are known to the reader. Given a deterministic finite-state automaton $A = (Q, s, \Sigma, \delta, F)$ over an ordered alphabet $(\Sigma, \prec)$, a natural partial order on its states can be defined in terms of the strings reaching them. For each state $q \in Q$, let $I_q = \{\alpha \in \Sigma^* \mid \delta(s, \alpha) = q\}$ denote the set of strings reaching q. The co-lexicographic partial order $<_A$ on the set of states of A is defined as follows: for states $p, q \in Q$, we set $p <_A q \iff \forall \alpha \in I_p \forall \beta \in I_q (\alpha \prec \beta)$. A *Wheeler automaton* is then defined as a deterministic finite-state automaton in which the order $<_A$ is a total order[2].

The class of Wheeler languages is defined accordingly: a language $\mathcal{L}$ is Wheeler iff there exists a Wheeler DFA that recognizes $\mathcal{L}$.

We next consider the transducer notion. A one-way *subsequential transducer* (which we simply call a transducer in this paper) computes a partial function from Σ^* to Γ^*, where Σ and Γ are the input and output alphabets, respectively. It is defined as a tuple $T = (Q, s, \Sigma, \Gamma, \delta, \omega, F, t)$, where:

1. $A_T = (Q, s, \Sigma, \delta, F)$ is the underlying DFA;
2. the partial function $\omega : Q \times \Sigma \to \Gamma^*$ has the same domain as δ;
3. $t : F \to \Gamma^*$ is the *terminal function*, which associates with each final state a string to be concatenated to the output produced while reading an accepted input string.

We call δ and ω the *input* and *output* transition functions, respectively, and we extend them as usual to $Q \times \Sigma^*$.

Given a transducer T, we denote by $||T|| : \Sigma^* \to \Gamma^*$ the partial function $\alpha \mapsto \omega(s, \alpha)t(\delta(s, \alpha))$, where $\alpha \in \Sigma^*$ is accepted by the underlying DFA A_T (see [6] for formal definitions).

Example 2.1. For the transducer in Fig. 1 we have $dom(||T||) = aab(ab)^*d + cab(ab)^*d$, $||T||(aab(ab)^kd) = 000(010)^k01$, $||T||(cab(ab)^kd) = 110(110)^k11$.

[1] see [7] for an in-depth discussion on the relevance of continuity for transducers
[2] There is also a notion of nondeterministic Wheeler automata, which we omit for space reasons

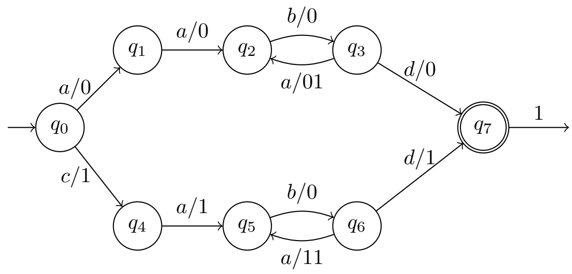

Figure 1: An example of a (Wheeler) transducer. The terminal function t is defined as: $t(q_7) = 1$.

3. Wheeler Transducers

We introduce the notion of deterministic finite-state Wheeler transducers and study their fundamental properties. We first show that the class is closed under composition, and we also prove that Wheeler languages are closed under inverse images of Wheeler transductions. We then turn to minimization: building on Choffrut's theory of minimal sequential transducers [8], we introduce a refinement of the syntactic Choffrut equivalence, denoted $\sim_f^c$, and show that it characterizes exactly the functions realizable by a Wheeler transducer. Finally, we provide a machine-independent characterization of functions computed by Wheeler transducers in terms of piecewise monotonicity [3].

We first present our proposal for the class of Wheeler transducers.

Definition 3.1. A transducer $T = (Q, s, \Sigma, \Gamma, \delta, \omega, F, t)$ is said to be a Wheeler transducer if, denoting by $<_{\text{in}}$ the input order $<_{A_T}$ on Q induced by the underlying DFA $A_T = (Q, s, \Sigma, \delta, F)$, the following conditions hold:

1. the order $<_{\text{in}}$ is total on Q;
2. for every state $q \in Q$, the function $\omega(s, -)$ is piecewise monotone on I_q.

A function f is said to be a *Wheeler function* if it is computed by a Wheeler transducer.

An example of a Wheeler transducer is shown in Fig. 1, where the co-lex order on states is $q_0 <_{\text{in}} q_1 <_{\text{in}} q_2 <_{\text{in}} q_5 <_{\text{in}} q_3 <_{\text{in}} q_6 <_{\text{in}} q_4 <_{\text{in}} q_7$ (a total order) and one can verify that condition 2 of Definition 3.1 is satisfied because, for each state q, the restriction of $\omega(s, -)$ to I_q is piecewise monotone. For instance, ω is monotone decreasing on I_{q_5}, hence piecewise monotone with one piece, while it is piecewise monotone with two pieces on I_{q_7}.

Our first result concerns the closure of Wheeler transducers under composition:

Lemma 3.2. *Let T_1, T_2 be two Wheeler transducers. Then the composition $\|T_2\| \circ \|T_1\|$ is realized by a Wheeler transducer.*

To support the claim that Wheeler transducers are well suited to representing the dynamics of Wheeler languages, our next result states that Wheeler languages are closed under inverse images of Wheeler transducers.

Lemma 3.3. *Suppose T is a Wheeler transducer and let $\mathcal{L} \subseteq \Gamma^*$ be a Wheeler language. Then the inverse image $||T||^{-1}(\mathcal{L}) \subseteq \Sigma^*$ is a Wheeler language.*

[3] A piecewise monotone function is a function in which the domain is split into a finite number of convex sets where the function is monotone (either non-decreasing or non-increasing)

3.1. Minimization

In the general case, given a partial function $f : \Sigma^* \to \Gamma^*$, Choffrut defined in [8] a syntactic congruence $\sim_f$ with the property that f is computed by a transducer if and only if $\sim_f$ has finite index. Moreover, the equivalence $\sim_f$ allows us to define a "minimum" transducer computing the function f.

In order to define the congruence $\sim_f$, one first defines a (partial) function $\widehat{f} : \Sigma^* \to \Gamma^*$ which, given a word α, outputs the longest common prefix of all words $f(\alpha\beta)$ for all continuations β with $\alpha\beta \in dom(f)$. More precisely, for $\alpha \in \mathrm{Pref}(dom(f))$, let

$$\widehat{f}(\alpha) = \bigwedge_{\alpha\beta \in dom(f)} f(\alpha\beta)$$

Then $\widehat{f}$ is used to define the equivalence $\sim_f$ over $\mathrm{Pref}(dom(f))$ as follows:

$$\alpha \sim_f \alpha' \Leftrightarrow \forall\beta \left(\widehat{f}(\alpha)^{-1} f(\alpha\beta) = \widehat{f}(\alpha')^{-1} f(\alpha'\beta)\right)$$

To obtain the corresponding syntactic congruence $\sim_f^c$ for Wheeler transducers, we need a convexity constraint on $\sim_f$, together with an order constraint on $\widehat{f}$. In the following, we denote by $[\alpha]_f$ the equivalence class of the word α with respect to $\sim_f$ and, for words $\mu, \mu' \in \mathrm{Pref}(dom(f))$, we denote by $[\mu, \mu']^{\pm}$ the following interval in $(\mathrm{Pref}(dom(f)), \prec)$: $[\mu, \mu']^{\pm} = \begin{cases} [\mu, \mu'] \text{ if } \mu \preceq \mu' \\ [\mu', \mu] \text{ if } \mu' \preceq \mu \end{cases}$

Definition 3.4. Let $f : \Sigma^* \to \Gamma^*$ be a partial function. Then the relation $\sim_f^c$ on $\mathrm{Pref}(dom(f))$ is defined as follows:

$$\alpha \sim_f^c \alpha' \iff [\alpha, \alpha']^{\pm} \subseteq [\alpha]_f \ \wedge\ \widehat{f} \text{ is piecewise monotone over } [\alpha, \alpha']^{\pm}$$

Lemma 3.5. *Given a partial function $f : \Sigma^* \to \Gamma^*$, the relation $\alpha \sim_f^c \alpha'$ is a right-invariant equivalence.*

The congruence $\sim_f^c$ allows us to establish a Myhill–Nerode theorem for Wheeler transducers.

Theorem 3.6. *A partial function $f : \Sigma^* \to \Gamma^*$ is computed by a Wheeler transducer if and only if $\sim_f^c$ has finite index and $\widehat{f}$ is piecewise monotone over each $\sim_f^c$-class.*

Using the previous theorem we may also obtain a characterization of Wheeler functions that highlights the monotonicity property directly on the function f, rather than on $\widehat{f}$:

Theorem 3.7. *A sequential function $f : \Sigma^* \to \Gamma^*$ is a Wheeler function if and only if every monotone sequence in $\mathrm{Pref}(dom(f))$ eventually lies in a single $\sim_f$-class and f is piecewise monotone.*

4. Conclusions and Future Work

The study initiated here paves the way for a broader research program on the structural theory of Wheeler transducers.

Several fundamental problems concerning decidability and complexity remain unresolved. The most immediate is the decidability of membership of a transducer in the class of Wheeler transducers and the complexity of such a decision procedure. A related problem is that of Wheeler realizability: given an arbitrary transducer, can one decide whether an equivalent Wheeler transducer exists? The complexity of the minimization procedure induced by $\sim_f^c$ likewise remains open.

A further direction for future work is to generalize the notion of Wheeler transducers beyond the present setting. In particular, it would be interesting to develop a hierarchy of transducers mirroring the hierarchy of regular languages induced by the deterministic width of their co-lexicographic order [9]. Such a hierarchy could provide a finer classification of transducers, relating their structural complexity to the degree to which their underlying automata depart from the Wheeler case. This may also lead to new algorithmic techniques for indexing and processing transductions with bounded co-lexicographic width.

Beyond the deterministic case, a natural extension is the study of Wheeler relations, analogous to rational relations in the regular setting, obtained by dropping the determinism requirement. In order to proceed in this direction, it is necessary to develop a theory of nondeterministic Wheeler transducers and to understand which properties of rational relations survive in the Wheeler setting. Finally, a logical characterization of Wheeler functions, analogous to the connection between regular functions and monadic second-order logic, is another direction for future work.

Acknowledgments

We thank the anonymous referee for a careful reading of the manuscript and for pointing out a serious flaw in a previous definition of $\sim_f^c$.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT to check grammar and spelling, rephrase and reword text, and create examples. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] J. Alanko, G. D'Agostino, A. Policriti, N. Prezza, Wheeler languages, Information and Computation (2021).

[2] T. C. P.-G. Consortium, Computational pan-genomics: status, promises and challenges, Briefings in Bioinformatics (2018). doi:10.1093/bib/bbw089.

[3] A. Backurs, P. Indyk, Which regular expression patterns are hard to match?, 2016. URL: https://arxiv.org/abs/1511.07070. arXiv:1511.07070.

[4] T. Gagie, G. Manzini, J. Sirén, Wheeler graphs: A framework for bwt-based data structures, Theoretical Computer Science 698 (2017). doi:10.1016/j.tcs.2017.06.016.

[5] E. Filiot, P.-A. Reynier, Transducers, logic and algebra for functions of finite words, ACM SIGLOG News (2016).

[6] A. Muscholl, G. Puppis, The Many Facets of String Transducers (Invited Talk), LIPIcs, Volume 126, STACS 2019 (2019).

[7] M. Bojańczyk, Transducers, 2026. URL: https://mimuw.edu.pl/~bojan/papers/transducer-book/transducer-book-2026-05-25.pdf, draft book, version of 25 May 2026.

[8] C. Choffrut, Minimizing subsequential transducers: a survey, Theoretical Computer Science 292 (2003). doi:10.1016/S0304-3975(01)00219-5.

[9] N. Cotumaccio, G. D'Agostino, A. Policriti, N. Prezza, Co-lexicographically ordering automata and regular languages - part i, Journal of the ACM (JACM) 70 (2023) 1–36.