

Some Results on Fixed Points of the Burrows-Wheeler Transform

Francesco DOLCE and Christian B. HUGHES



Czech Technical University in Prague

Naples, Italy 2026



Cyclic Conjugates and Sorting

Cyclic Conjugates and Sorting

Let $w = \text{vesuvio}$ over the ordered Italian alphabet $\mathcal{I} = \{a < b < \dots < z\}$.

Cyclic Conjugates and Sorting

Let $w = \text{vesuvio}$ over the ordered Italian alphabet $\mathcal{I} = \{a < b < \dots < z\}$.

`vesuvio`

Cyclic Conjugates and Sorting

Let $w = \text{vesuvio}$ over the ordered Italian alphabet $\mathcal{I} = \{a < b < \dots < z\}$.

vesuvio

esuviov

Cyclic Conjugates and Sorting

Let $w = \text{vesuvio}$ over the ordered Italian alphabet $\mathcal{I} = \{a < b < \dots < z\}$.

vesuvio

esuviov

suviove

Cyclic Conjugates and Sorting

Let $w = \text{vesuvio}$ over the ordered Italian alphabet $\mathcal{I} = \{a < b < \dots < z\}$.

vesuvio

esuviov

suviove

uvioves

Cyclic Conjugates and Sorting

Let $w = \text{vesuvio}$ over the ordered Italian alphabet $\mathcal{I} = \{a < b < \dots < z\}$.

vesuvio

esuviov

suviove

uvioves

viovesu

Cyclic Conjugates and Sorting

Let $w = \text{vesuvio}$ over the ordered Italian alphabet $\mathcal{I} = \{a < b < \dots < z\}$.

vesuvio

esuviov

suviove

uvioves

viovesu

iovesuv

Cyclic Conjugates and Sorting

Let $w = \text{vesuvio}$ over the ordered Italian alphabet $\mathcal{I} = \{a < b < \dots < z\}$.

vesuvio

esuviov

suviove

uvioves

viovesu

iovesuv

ovesuvi

Cyclic Conjugates and Sorting

Let $w = \text{vesuvio}$ over the ordered Italian alphabet $\mathcal{I} = \{a < b < \dots < z\}$.

esuviov

iovesuv

ovesuvi

suviove

uvioves

vesuvio

viovesu

Cyclic Conjugates and Sorting

Let $w = \text{vesuvio}$ over the ordered Italian alphabet $\mathcal{I} = \{a < b < \dots < z\}$.

$$\begin{bmatrix} \text{esuvi ov} \\ \text{iovesu v} \\ \text{ovesu vi} \\ \text{suvio ve} \\ \text{uvio ves} \\ \text{vesu v io} \\ \text{v io vesu} \end{bmatrix}$$

The Burrows-Wheeler Transform

$$\begin{bmatrix} \text{esuviov} \\ \text{iovesuv} \\ \text{ovesuvi} \\ \text{suviove} \\ \text{uvioves} \\ \text{vesuvio} \\ \text{viovesu} \end{bmatrix}$$

The Burrows-Wheeler Transform

$$\begin{bmatrix} \text{esu} \text{vio} \textbf{v} \\ \text{iovesu} \textbf{v} \\ \text{ovesu} \textbf{i} \\ \text{suvio} \textbf{e} \\ \text{uvio} \textbf{es} \\ \text{vesu} \textbf{io} \\ \text{vio} \textbf{es} \textbf{u} \end{bmatrix}$$

The Burrows-Wheeler Transform

$$\begin{bmatrix} \text{esu} \mathbf{v} \text{io} \mathbf{v} \\ \text{io} \mathbf{v} \text{esu} \mathbf{v} \\ \text{ov} \mathbf{e} \text{su} \mathbf{i} \\ \text{su} \mathbf{v} \text{io} \mathbf{v} \mathbf{e} \\ \text{u} \mathbf{v} \text{io} \mathbf{v} \mathbf{e} \mathbf{s} \\ \text{v} \mathbf{e} \text{su} \mathbf{v} \text{io} \\ \text{v} \text{io} \mathbf{v} \mathbf{e} \mathbf{s} \mathbf{u} \end{bmatrix}$$

$$\text{bwt}_{\mathcal{I}}(\text{vesuvio}) = \mathbf{v} \mathbf{v} \mathbf{i} \mathbf{e} \mathbf{s} \mathbf{o} \mathbf{u}$$

Runs, Clustering, and Δ

Runs, Clustering, and Δ

A *run* of w is a maximal factor a^k with $a \in \mathcal{A}$, $k \geq 1$. We denote the number of runs in w by $\rho(w)$.

$$\text{bwt}_{\mathcal{I}}(\text{vesuvio}) = \text{vv} \mid \text{i} \mid \text{e} \mid \text{s} \mid \text{o} \mid \text{u}, \quad \rho = 6, \quad |\text{alph}(\text{bwt}_{\mathcal{I}}(w))| = 6.$$

Runs, Clustering, and Δ

A *run* of w is a maximal factor a^k with $a \in \mathcal{A}$, $k \geq 1$. We denote the number of runs in w by $\rho(w)$.

$$\text{bwt}_{\mathcal{I}}(\text{vesuvio}) = \text{vv} \mid \text{i} \mid \text{e} \mid \text{s} \mid \text{o} \mid \text{u}, \quad \rho = 6, \quad |\text{alph}(\text{bwt}_{\mathcal{I}}(w))| = 6.$$

Every letter has at least one run, so

$$\Delta(w) := \rho(w) - |\text{alph}(w)| \geq 0,$$

Runs, Clustering, and Δ

A *run* of w is a maximal factor a^k with $a \in \mathcal{A}$, $k \geq 1$. We denote the number of runs in w by $\rho(w)$.

$$\text{bwt}_{\mathcal{I}}(\text{vesuvio}) = \text{vv} | \text{i} | \text{e} | \text{s} | \text{o} | \text{u}, \quad \rho = 6, \quad |\text{alph}(\text{bwt}_{\mathcal{I}}(w))| = 6.$$

Every letter has at least one run, so

$$\Delta(w) := \rho(w) - |\text{alph}(w)| \geq 0,$$

We see that $\Delta(\text{vesuvio}) = 7 - 6 = 1$.

Runs, Clustering, and Δ

A *run* of w is a maximal factor a^k with $a \in \mathcal{A}$, $k \geq 1$. We denote the number of runs in w by $\rho(w)$.

$$\text{bwt}_{\mathcal{I}}(\text{vesuvio}) = \text{vv} \mid \text{i} \mid \text{e} \mid \text{s} \mid \text{o} \mid \text{u}, \quad \rho = 6, \quad |\text{alph}(\text{bwt}_{\mathcal{I}}(w))| = 6.$$

Every letter has at least one run, so

$$\Delta(w) := \rho(w) - |\text{alph}(w)| \geq 0,$$

We see that $\Delta(\text{vesuvio}) = 7 - 6 = 1$.

A word w is *π -clustering* if $\text{bwt}_{\mathcal{A}}(w) = a_{\pi(a_1)}^{k_1} a_{\pi(a_2)}^{k_2} \cdots a_{\pi(a_n)}^{k_n}$. The word $w = \text{vesuvio}$ clusters with $\pi = (6, 2, 1, 4, 3, 5)$ when considered as a word over $\mathcal{A} = \{e < i < o < s < u < v\}$.

Runs, Clustering, and Δ

A **run** of w is a maximal factor a^k with $a \in \mathcal{A}$, $k \geq 1$. We denote the number of runs in w by $\rho(w)$.

$$\text{bwt}_{\mathcal{I}}(\text{vesuvio}) = \text{vv} \mid \text{i} \mid \text{e} \mid \text{s} \mid \text{o} \mid \text{u}, \quad \rho = 6, \quad |\text{alph}(\text{bwt}_{\mathcal{I}}(w))| = 6.$$

Every letter has at least one run, so

$$\Delta(w) := \rho(w) - |\text{alph}(w)| \geq 0,$$

We see that $\Delta(\text{vesuvio}) = 7 - 6 = 1$.

A word w is **π -clustering** if $\text{bwt}_{\mathcal{A}}(w) = a_{\pi(a_1)}^{k_1} a_{\pi(a_2)}^{k_2} \cdots a_{\pi(a_n)}^{k_n}$. The word $w = \text{vesuvio}$ clusters with $\pi = (6, 2, 1, 4, 3, 5)$ when considered as a word over $\mathcal{A} = \{e < i < o < s < u < v\}$.

Theorem (Mantaci, Restivo, Rosone, Sciortino, Versari 2017)

$$|\text{alph}(w)| \leq \rho(\text{bwt}_{\mathcal{A}}(w)) \leq 2\rho(w)$$

Which Words are BWT Images ?

Which Words are BWT Images ?

The BWT is not surjective (e.g., $w = ab$). A sequence of letters $(b_1, b_2, \dots, b_n)$ has a *global ascent* if

$$\max_{1 \leq i \leq h} b_i \leq \min_{h < j \leq n} b_j \quad \text{for some } 1 \leq h < n.$$

Which Words are BWT Images?

The BWT is not surjective (e.g., $w = ab$). A sequence of letters $(b_1, b_2, \dots, b_n)$ has a *global ascent* if

$$\max_{1 \leq i \leq h} b_i \leq \min_{h < j \leq n} b_j \quad \text{for some } 1 \leq h < n.$$

Theorem (Likhomanov, Shur 2011)

$b_1^{k_1} b_2^{k_2} \dots b_n^{k_n}$, for some $k_i > 0$, $b_i \neq b_{i+1}$, is an image of the BWT if and only if $(b_1, \dots, b_n)$ has no global ascent.

Which Words are BWT Images ?

The BWT is not surjective (e.g., $w = ab$). A sequence of letters $(b_1, b_2, \dots, b_n)$ has a *global ascent* if

$$\max_{1 \leq i \leq h} b_i \leq \min_{h < j \leq n} b_j \quad \text{for some } 1 \leq h < n.$$

Theorem (Likhomanov, Shur 2011)

$b_1^{k_1} b_2^{k_2} \dots b_n^{k_n}$, for some $k_i > 0$, $b_i \neq b_{i+1}$, is an image of the BWT if and only if $(b_1, \dots, b_n)$ has no global ascent.

The word $\text{bwt}_{\mathcal{I}}(\text{vesuvio}) = \text{vviesou}$ has run sequence (v, i, e, s, o, u) , and is thus the image of some word under the BWT. The word eiosuvv , however, is not a BWT image of any word.

Which Words are BWT Images ?

The BWT is not surjective (e.g., $w = ab$). A sequence of letters $(b_1, b_2, \dots, b_n)$ has a *global ascent* if

$$\max_{1 \leq i \leq h} b_i \leq \min_{h < j \leq n} b_j \quad \text{for some } 1 \leq h < n.$$

Theorem (Likhomanov, Shur 2011)

$b_1^{k_1} b_2^{k_2} \dots b_n^{k_n}$, for some $k_i > 0$, $b_i \neq b_{i+1}$, is an image of the BWT if and only if $(b_1, \dots, b_n)$ has no global ascent.

The word $\text{bwt}_{\mathcal{I}}(\text{vesuvio}) = \text{vviesou}$ has run sequence (v, i, e, s, o, u) , and is thus the image of some word under the BWT. The word eiosuvv , however, is not a BWT image of any word.

Corollary

If an image of the BWT is a Lyndon word, then it is a single letter.

Fixed Points

Fixed Points

A word $w \in \mathcal{A}^*$ is a *fixed point* of the BWT if $\text{bwt}_{\mathcal{A}}(w) = w$. Sorting the conjugates of $w = \text{pino}$ over $\mathcal{I} = \{i < n < o < p\}$ gives

$\text{inop} < \text{nopi} < \text{opin} < \text{pino}$.

Fixed Points

A word $w \in \mathcal{A}^*$ is a *fixed point* of the BWT if $\text{bwt}_{\mathcal{A}}(w) = w$. Sorting the conjugates of $w = \text{pino}$ over $\mathcal{I} = \{i < n < o < p\}$ gives

$$\text{inop} < \text{nopi} < \text{opin} < \text{pino}.$$

Corollary

The only fixed points of the BWT that are Lyndon words are the single letters.

Fixed Points

A word $w \in \mathcal{A}^*$ is a *fixed point* of the BWT if $\text{bwt}_{\mathcal{A}}(w) = w$. Sorting the conjugates of $w = \text{pino}$ over $\mathcal{I} = \{i < n < o < p\}$ gives

$$\text{inop} < \text{nopi} < \text{opin} < \text{pino}.$$

Corollary

The only fixed points of the BWT that are Lyndon words are the single letters.

What other explicit families of fixed points can we give?

Fixed Points

A word $w \in \mathcal{A}^*$ is a *fixed point* of the BWT if $\text{bwt}_{\mathcal{A}}(w) = w$. Sorting the conjugates of $w = \text{pino}$ over $\mathcal{I} = \{i < n < o < p\}$ gives

$$\text{inop} < \text{nop} < \text{opin} < \text{pino}.$$

Corollary

The only fixed points of the BWT that are Lyndon words are the single letters.

What other explicit families of fixed points can we give?

The Plan (for w with $|\text{alph}(w)| > 2$)

$\Delta(w) = 0$	$\Delta(w) = 1$	$\Delta(w) \geq 2$
characterized	characterized	open

Non-Primitive Fixed Points

Non-Primitive Fixed Points

A word w is *primitive* if $w = u^p$ implies $p = 1$. We have

$$\text{bwt}_{\mathcal{A}}(u) = b_1 b_2 \cdots b_n \implies \text{bwt}_{\mathcal{A}}(u^p) = b_1^p b_2^p \cdots b_n^p.$$

Non-Primitive Fixed Points

A word w is *primitive* if $w = u^p$ implies $p = 1$. We have

$$\text{bwt}_{\mathcal{A}}(u) = b_1 b_2 \cdots b_n \implies \text{bwt}_{\mathcal{A}}(u^p) = b_1^p b_2^p \cdots b_n^p.$$

So u^p is a fixed point exactly when $b_1^p b_2^p \cdots b_n^p = u^p$.

Non-Primitive Fixed Points

A word w is *primitive* if $w = u^p$ implies $p = 1$. We have

$$\text{bwt}_{\mathcal{A}}(u) = b_1 b_2 \cdots b_n \implies \text{bwt}_{\mathcal{A}}(u^p) = b_1^p b_2^p \cdots b_n^p.$$

So u^p is a fixed point exactly when $b_1^p b_2^p \cdots b_n^p = u^p$.

Proposition (Dolce, H. 2026+)

If $w \in \mathcal{A}^+$ is a fixed point of the BWT and $w = u^p$ with $p > 1$, then either $w = a^p$ for a single letter a , or $w = (b^2 a^2)^2$ for $a < b$.

Non-Primitive Fixed Points

A word w is *primitive* if $w = u^p$ implies $p = 1$. We have

$$\text{bwt}_{\mathcal{A}}(u) = b_1 b_2 \cdots b_n \implies \text{bwt}_{\mathcal{A}}(u^p) = b_1^p b_2^p \cdots b_n^p.$$

So u^p is a fixed point exactly when $b_1^p b_2^p \cdots b_n^p = u^p$.

Proposition (Dolce, H. 2026+)

If $w \in \mathcal{A}^+$ is a fixed point of the BWT and $w = u^p$ with $p > 1$, then either $w = a^p$ for a single letter a , or $w = (b^2 a^2)^2$ for $a < b$.

In particular, $|\text{alph}(w)| > 2$ implies w is primitive, so every fixed point onward is primitive.

$\Delta = 0$: *Clustering Fixed Points*

$\Delta = 0$: *Clustering Fixed Points*

Imposing $\Delta(w) = 0$ for a fixed point w forces w to have exactly one run per letter.
Among fixed points, $\Delta = 0$ and clustering are the same condition.

$\Delta = 0$: *Clustering Fixed Points*

Imposing $\Delta(w) = 0$ for a fixed point w forces w to have exactly one run per letter.
Among fixed points, $\Delta = 0$ and clustering are the same condition.

Theorem (Dolce, H. 2026+)

Let w be a word with $|\text{alph}(w)| > 2$. Then w is a fixed point of the BWT with $\Delta(w) = 0$ if and only if $w = a_m a_1^{k_1} a_2^{k_2} \dots a_{m-1}^{k_{m-1}}$, for some $a_1 < a_2 < \dots < a_m$ in $\mathcal{A}$ and integers $k_i > 0$.

$\Delta = 0$: *Clustering Fixed Points*

Imposing $\Delta(w) = 0$ for a fixed point w forces w to have exactly one run per letter. Among fixed points, $\Delta = 0$ and clustering are the same condition.

Theorem (Dolce, H. 2026+)

Let w be a word with $|\text{alph}(w)| > 2$. Then w is a fixed point of the BWT with $\Delta(w) = 0$ if and only if $w = a_m a_1^{k_1} a_2^{k_2} \dots a_{m-1}^{k_{m-1}}$, for some $a_1 < a_2 < \dots < a_m$ in $\mathcal{A}$ and integers $k_i > 0$.

The largest letter occurs once, followed by other smaller letters in increasing order. For instance, $w = \text{eaabcbdd}$ over $\{a < b < c < d < e\}$.

$\Delta = 0$: *Clustering Fixed Points*

Imposing $\Delta(w) = 0$ for a fixed point w forces w to have exactly one run per letter. Among fixed points, $\Delta = 0$ and clustering are the same condition.

Theorem (Dolce, H. 2026+)

Let w be a word with $|\text{alph}(w)| > 2$. Then w is a fixed point of the BWT with $\Delta(w) = 0$ if and only if $w = a_m a_1^{k_1} a_2^{k_2} \dots a_{m-1}^{k_{m-1}}$, for some $a_1 < a_2 < \dots < a_m$ in $\mathcal{A}$ and integers $k_i > 0$.

The largest letter occurs once, followed by other smaller letters in increasing order. For instance, $w = \text{eaabccdd}$ over $\{a < b < c < d < e\}$.

Note that every clustering fixed point is maximal in its conjugacy class.

$\Delta = 1$: *Ternary Case*

$\Delta = 1$: *Ternary Case*

We now consider the case when exactly one letter has two runs. Over $\mathcal{A} = \{a < b < c\}$ a fixed point with $\Delta(w) = 1$ has four runs.

$\Delta = 1$: Ternary Case

We now consider the case when exactly one letter has two runs. Over $\mathcal{A} = \{a < b < c\}$ a fixed point with $\Delta(w) = 1$ has four runs.

Theorem (Dolce, H. 2026+)

A word $w \in \{a < b < c\}^*$ is a fixed point of the BWT with $\Delta(w) = 1$ if and only if

$$w = c^{2k+1} a c^k b, \quad w = c^{2k} b c^k a, \quad \text{or} \quad w = c^2 b^k c a,$$

for some integer $k > 0$.

$\Delta = 1$: Ternary Case

We now consider the case when exactly one letter has two runs. Over $\mathcal{A} = \{a < b < c\}$ a fixed point with $\Delta(w) = 1$ has four runs.

Theorem (Dolce, H. 2026+)

A word $w \in \{a < b < c\}^*$ is a fixed point of the BWT with $\Delta(w) = 1$ if and only if

$$w = c^{2k+1} a c^k b, \quad w = c^{2k} b c^k a, \quad \text{or} \quad w = c^2 b^k c a,$$

for some integer $k > 0$.

The word $c^2 b^k c a$ has an immediate analogue on larger alphabets.

$\Delta = 1$: *Four or More Letters*

$\Delta = 1$: *Four or More Letters*

Theorem (Dolce, H. 2026+)

Let w be a fixed point of the BWT with $|\text{alph}(w)| > 3$ and $\Delta(w) = 1$. Then $w = a_m^2 a_2^{k_2} a_3^{k_3} \dots a_{m-1}^{k_{m-1}} a_m a_1$ for some $a_1 < a_2 < \dots < a_m \in \mathcal{A}$ and positive integers k_i .

$\Delta = 1$: *Four or More Letters*

Theorem (Dolce, H. 2026+)

Let w be a fixed point of the BWT with $|\text{alph}(w)| > 3$ and $\Delta(w) = 1$. Then $w = a_m^2 a_2^{k_2} a_3^{k_3} \dots a_{m-1}^{k_{m-1}} a_m a_1$ for some $a_1 < a_2 < \dots < a_m \in \mathcal{A}$ and positive integers k_i .

Examples :

$\Delta = 1$: *Four or More Letters*

Theorem (Dolce, H. 2026+)

Let w be a fixed point of the BWT with $|\text{alph}(w)| > 3$ and $\Delta(w) = 1$. Then $w = a_m^2 a_2^{k_2} a_3^{k_3} \dots a_{m-1}^{k_{m-1}} a_m a_1$ for some $a_1 < a_2 < \dots < a_m \in \mathcal{A}$ and positive integers k_i .

Examples :

- eebbbcdea

$\Delta = 1$: *Four or More Letters*

Theorem (Dolce, H. 2026+)

Let w be a fixed point of the BWT with $|\text{alph}(w)| > 3$ and $\Delta(w) = 1$. Then $w = a_m^2 a_2^{k_2} a_3^{k_3} \dots a_{m-1}^{k_{m-1}} a_m a_1$ for some $a_1 < a_2 < \dots < a_m \in \mathcal{A}$ and positive integers k_i .

Examples :

- eebbbcdea
- ddbcda

$\Delta = 1$: *Four or More Letters*

Theorem (Dolce, H. 2026+)

Let w be a fixed point of the BWT with $|\text{alph}(w)| > 3$ and $\Delta(w) = 1$. Then $w = a_m^2 a_2^{k_2} a_3^{k_3} \dots a_{m-1}^{k_{m-1}} a_m a_1$ for some $a_1 < a_2 < \dots < a_m \in \mathcal{A}$ and positive integers k_i .

Examples :

- eebbbcdea
- ddbcd a
- ddbccda

$\Delta = 1$: *Four or More Letters*

Theorem (Dolce, H. 2026+)

Let w be a fixed point of the BWT with $|\text{alph}(w)| > 3$ and $\Delta(w) = 1$. Then $w = a_m^2 a_2^{k_2} a_3^{k_3} \dots a_{m-1}^{k_{m-1}} a_m a_1$ for some $a_1 < a_2 < \dots < a_m \in \mathcal{A}$ and positive integers k_i .

Examples :

- eebbbcdea
- ddbcd a
- ddbccda
- ddbcccccccccccccccccccccccccccccccccccccd a

$$\Delta \geq 2 ?$$

$$\Delta \geq 2 ?$$

Proposition

The word $w = a_{m-1}^2 a_1^2 a_m a_2^+ a_3^+ \cdots a_{m-2}^+ a_{m-1} a_1^2$ over $\mathcal{A} = \{a_1 < \dots < a_m\}$ is a fixed point of the BWT when $m \geq 3$.

$$\Delta \geq 2 ?$$

Proposition

The word $w = a_{m-1}^2 a_1^2 a_m a_2^+ a_3^+ \cdots a_{m-2}^+ a_{m-1} a_1^2$ over $\mathcal{A} = \{a_1 < \dots < a_m\}$ is a fixed point of the BWT when $m \geq 3$.

We do not have a full list of families for any $\Delta \geq 2$.

The BWT and IET Languages

(More on this relationship tomorrow)

The BWT and IET Languages

(More on this relationship tomorrow)

Corollary (Dolce, H. 2026+)

Let T be an IET over the alphabet $\mathcal{A} = \{a_1 < \dots < a_n\}$. If there exists a pangrammatic fixed point of the BWT in the set of return words in $\mathcal{L}(T)$, then the permutation associated with T is $\pi = (a_1 a_n a_{n-1} \dots a_2)$.

The BWT and IET Languages

(More on this relationship tomorrow)

Corollary (Dolce, H. 2026+)

Let T be an IET over the alphabet $\mathcal{A} = \{a_1 < \dots < a_n\}$. If there exists a pangrammatic fixed point of the BWT in the set of return words in $\mathcal{L}(T)$, then the permutation associated with T is $\pi = (a_1 a_n a_{n-1} \dots a_2)$.

Corollary (Dolce, H. 2026+)

Let T be a symmetric IET over the alphabet $\mathcal{A} = \{a_1 < \dots < a_n\}$. Every fixed point in the set of return words in $\mathcal{L}(T)$ contains at most two letters.

Grazie mille per l'attenzione !