# On Block Pumpable Languages

**Christopher Hanrui Chak, Singapore**

**Rusins Freivalds, Riga (1942 - 2016)**

**Frank Stephan, Singapore**

**Henrietta Tan Wan Yik, Singapore**

# Finite Automata

## Recognising Multiples of Three

Three states: Remainders $\mathbf{0}$ (initial), $\mathbf{1}$, $\mathbf{2}$.
Update of state on digit: $(\mathbf{s}, \mathbf{d}) \mapsto (\mathbf{s} + \mathbf{d}) \bmod \mathbf{3}$;
for example, state $\mathbf{2}$ and input $\mathbf{8}$ give new state $\mathbf{1}$.
Accept numbers where final state is $\mathbf{0}$.

```
Input:   2 5 6 1 0 2 4 2 0 4 8
State: 0 2 1 1 2 2 1 2 1 1 2 1
Final Decision: Reject
```

## Multiples of $\mathbf{p}$

States $\{\mathbf{0}, \mathbf{1}, \ldots, \mathbf{p} - \mathbf{1}\}$; initial state $\mathbf{0}$.
Update: $(\mathbf{s}, \mathbf{d}) \mapsto ((\mathbf{s} \cdot \mathbf{10}) + \mathbf{d}) \bmod \mathbf{p}$.
Accept numbers where final state is $\mathbf{0}$.

# Pumping

Given a word $\mathbf{x}$, consider two breakpoints $(\mathbf{i})$, $(\mathbf{j})$ which split the word three parts: $\mathbf{x} = \mathbf{u}(\mathbf{i})\mathbf{v}(\mathbf{j})\mathbf{w}$. Now consider words which repeat or omit $\mathbf{v}$:

$$\mathbf{P^k(x : i, j) = u\,v^k\,w},$$
$$\mathbf{P^*(x : i, j) = \{P^k(x : i, j) : k \geq 0\}},$$
$$\mathbf{P^+(x : i, j) = \{P^k(x : i, j) : k > 0\}}.$$

Traditional Pumping Lemma [Scott and Rabin 1959; Bar-Hillel, Perles and Shamir 1961]
If a language is regular then there is a constant $\mathbf{c}$ such that for every $\mathbf{x} \in \mathbf{L}$ longer than $\mathbf{c}$ there are breakpoints $(\mathbf{i}), (\mathbf{j})$ within the first $\mathbf{c}$ symbols of $\mathbf{x}$ such that $\mathbf{P^*(x : i, j) \subseteq L}$.

Example: if $\mathbf{L} = \{\mathbf{00}\}^* \cdot \{\varepsilon, \mathbf{01}\} \cdot \{\mathbf{11}\}^*$ then $\mathbf{c} = \mathbf{4}$ and either the first two or the second two symbols can be pumped: $\mathbf{x} = (\mathbf{i})\mathbf{00}(\mathbf{j})\mathbf{01}$ or $\mathbf{x} = \mathbf{01}(\mathbf{i})\mathbf{11}(\mathbf{j})\mathbf{11}$ satisfy $\mathbf{P^*(x : i, j) \subseteq L}$.

# Matching Pumping Lemmas

Jaffe's Pumping Lemma [Jaffe 1978]
A language $\mathbf{L} \subseteq \Sigma^*$ is regular iff there is a constant $\mathbf{c}$ such that for all $\mathbf{x} \in \Sigma^*$ and $\mathbf{y} \in \Sigma^{\mathbf{c}}$ there are two breakpoints splitting $\mathbf{y}$ into $\mathbf{u}, \mathbf{v}, \mathbf{w}$ with $\mathbf{v} \neq \varepsilon$ and, for all $\mathbf{h} \in \mathbb{N}$, $\mathbf{L}_{\mathbf{x}\mathbf{u}\mathbf{v}^{\mathbf{h}}\mathbf{w}} = \mathbf{L}_{\mathbf{x}\mathbf{y}}$.

Here the derivative is defined as $\mathbf{L}_{\mathbf{z}} = \{\mathbf{z}' : \mathbf{z}\mathbf{z}' \in \mathbf{L}\}$.

This follows from Myhill and Nerode's Theorem that $\mathbf{L}$ is regular iff it has finitely many derivatives.

Block Pumping Lemma [Ehrenfeucht, Parikh and Rozenberg 1981]
$\mathbf{L}$ is regular iff there is a constant $\mathbf{c}$ such that for every word $\mathbf{x}$ with a set $\mathbf{K}$ of $\mathbf{c}$ breakpoints there are $\mathbf{i}, \mathbf{j} \in \mathbf{K}$ such that if $\mathbf{x} \in \mathbf{L}$ then $\mathbf{P}^*(\mathbf{x} : \mathbf{i}, \mathbf{j}) \subseteq \mathbf{L}$ else $\mathbf{P}^*(\mathbf{x} : \mathbf{i}, \mathbf{j}) \subseteq \overline{\mathbf{L}}$.

# Variants

Block Cancellation Lemma [Ehrenfeucht, Parikh and Rozenberg 1981]
$\mathbf{L}$ is regular iff there is a constant $\mathbf{c}$ such that for every word $\mathbf{x}$ with a set $\mathbf{K}$ of $\mathbf{c}$ breakpoints there are $\mathbf{i}, \mathbf{j} \in \mathbf{K}$ such that if $\mathbf{x} \in \mathbf{L}$ then $\mathbf{P^0(x : i, j)} \in \mathbf{L}$ else $\mathbf{P^0(x : i, j)} \notin \overline{\mathbf{L}}$.

They left the other variant – to look at $\mathbf{P^+(x : i, j)}$ – as an open problem which was solved by Varricchio sixteen years later.

Positive Block Pumping Lemma [Varricchio 1997]
$\mathbf{L}$ is regular iff there is a constant $\mathbf{c}$ such that for every word $\mathbf{x}$ with a set $\mathbf{K}$ of $\mathbf{c}$ breakpoints there are $\mathbf{i}, \mathbf{j} \in \mathbf{K}$ such that if $\mathbf{x} \in \mathbf{L}$ then $\mathbf{P^+(x : i, j)} \subseteq \mathbf{L}$ else $\mathbf{P^+(x : i, j)} \subseteq \overline{\mathbf{L}}$.

# Block Pumpable Languages

Definition [Chak, Freivalds, Stephan and Tan 2016]
A language $\mathbf{L}$ is block pumpable / block cancellable / positively block pumpable iff there is a constant $\mathbf{c}$ such that for every word $\mathbf{x} \in \mathbf{L}$ and set $\mathbf{K}$ of $\mathbf{c}$ breakpoints there are $\mathbf{i}, \mathbf{j} \in \mathbf{K}$ with $\mathbf{P}^*(\mathbf{x} : \mathbf{i}, \mathbf{j}) \subseteq \mathbf{L}$ / $\mathbf{P}^\mathbf{0}(\mathbf{x} : \mathbf{i}, \mathbf{j}) \in \mathbf{L}$ / $\mathbf{P}^+(\mathbf{x} : \mathbf{i}, \mathbf{j}) \subseteq \mathbf{L}$, respectively.

Examples
Let $\mathbf{L_{sc}}$ be the language $\mathbf{L}$ of all square containing words, that is, $\mathbf{L_{sc}} = \{\mathbf{uvvw} : \mathbf{v} \neq \varepsilon\}$. Then $\mathbf{L}$ is positively block pumpable with constant $\mathbf{c} = \mathbf{2}$; when the alphabet has at least three letters, $\mathbf{L_{sc}}$ is not regular.
The language $\mathbf{L_{exp}} = \{\mathbf{w} : |\mathbf{w}|$ is not a power of $\mathbf{3}\}$ is block cancellable with constant $\mathbf{4}$.
$\mathbf{L_{sc}} \cup \mathbf{L_{exp}}$ is block pumpable with constant $\mathbf{4}$ and not regular for any alphabet with at least three letters.

# Equivalence

Theorem [Chak, Freivalds, Stephan and Tan 2016]
A language is block pumpable iff it is both, block cancellable and positively block pumpable.

Proof for $(\Leftarrow)$: Assume that $\mathbf{L}$ is block cancellable and positively block pumpable with constant $\mathbf{c}$. Let $\mathbf{c}'$ be so large that if one colours the pairs of a set of $\mathbf{c}'$ elements with two colours then this set has a monochromatic subset with at least $\mathbf{c}$ elements.

Given a word $\mathbf{x} \in \mathbf{L}$ and a set $\mathbf{K}'$ of $\mathbf{c}'$ breakpoints, it has a subset of $\mathbf{K}$ of $\mathbf{c}$ breakpoints such that either, for all $\mathbf{i}, \mathbf{j} \in \mathbf{K}$, $\mathbf{P^0(x : i, j) \in L}$ or, for all $\mathbf{i}, \mathbf{j} \in \mathbf{K}$, $\mathbf{P^0(x : i, j) \notin L}$. As $\mathbf{L}$ is block cancellable with $\mathbf{c}$, the first happens.

Now $\mathbf{K}$ contains a pair $\mathbf{i}, \mathbf{j}$ of breakpoints such that $\mathbf{P^+(x : i, j) \subseteq L}$. As $\mathbf{P^0(x : i, j) \in L}$ as well, $\mathbf{P^*(x : i, j) \subseteq L}$. Thus $\mathbf{L}$ is block pumpable with constant $\mathbf{c}'$.

# Uses of Ramsey Theorem

Ramsey's Theorem of Pairs can also be used to prove the following results.

Theorem [Chak, Freivalds, Stephan and Tan 2016]
If $\mathbf{L}, \mathbf{H}$ are block pumpable then $\mathbf{L} \cap \mathbf{H}$ is also block pumpable.

Theorem [Chak, Freivalds, Stephan and Tan 2016]
If $\mathbf{L}$ is block pumpable and $\mathbf{f}$ computed by a finite transducer then $\mathbf{f}(\mathbf{L})$ is block pumpable.

Other Properties
The union of two block pumpable languages and the concatenation of two block pumpable languages are again block pumpable.

Similarly for block cancellable and positively block pumpable.

# Kleene Star

Example [Chak, Freivalds, Stephan and Tan 2016]
Let $\mathbf{L} \subseteq \{\mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}\}^*$ contain all words of the form $\mathbf{3v4w}$ such that $\mathbf{v}, \mathbf{w} \in \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}^*$ and at least one of the following conditions hold:

- $\mathbf{v} = \mathbf{w}$;
- $|\mathbf{v}| \neq |\mathbf{w}|$;
- $\mathbf{v}$ or $\mathbf{w}$ contains a square.

The language $\mathbf{L}$ is block pumpable with block pumping constant $\mathbf{6}$ and that $\mathbf{L}^*$ is neither block cancellable nor positively block pumpable.

# Proof

$\mathbf{L}$: If $\mathbf{x = x_0(1)x_1(2)x_2(3)x_3(4)x_4(5)x_5(6)x_6}$ is a given word in $\mathbf{L}$ with six breakpoints then there is a $\mathbf{k \in \{1, 2, 3, 4\}}$ such that $\mathbf{x_k, x_{k+1} \in \{0, 1, 2\}^+}$.

Now $\mathbf{P^+(x : k, k+1) \cup P^+(x : k, k+2) \subseteq L}$, due to $\mathbf{v}$ or $\mathbf{w}$ not square-free after pumping. Furthermore, $\mathbf{P^0(x : k, k+1) \in L}$ or $\mathbf{P^0(x : k, k+2) \in L}$, due to $\mathbf{|v| \neq |w|}$ after omitting.

$\mathbf{L^*}$: If $\mathbf{L^*}$ would be block cancellable or positively block pumpable with constant $\mathbf{c}$ then consider $\mathbf{c}$ distinct square-free words $\mathbf{u_1, u_2, \ldots, u_c \in \{0, 1, 2\}^*}$ of the same length and consider $\mathbf{3u_1(1)4u_1 3u_2(2)4u_2 \ldots 3u_c(c)4u_c}$. After pumping or omitting there is a subword $\mathbf{3u_i4u_j}$ with $\mathbf{i \neq j}$, while all the subwords $\mathbf{3u_i4u_j}$ satisfy $\mathbf{|u_i| = |u_j|}$ and $\mathbf{u_i, u_j}$ are square-free. Thus this word neither admits cancellation nor positive pumping at indicated breakpoints.

# Double Pumping

(1) If $\mathbf{L}$ is regular then there is a constant $\mathbf{c}$ such that for all $\mathbf{x} \in \mathbf{L}$ and for all sets $\mathbf{K}$ of $\mathbf{c}$ breakpoints on $\mathbf{x}$, there are breakpoints $\mathbf{i}, \mathbf{j}, \mathbf{k} \in \mathbf{K}$ splitting $\mathbf{x}$ into $\mathbf{u}, \mathbf{v}, \tilde{\mathbf{v}}, \mathbf{w}$ with $\mathbf{u}\mathbf{v}^*\tilde{\mathbf{v}}^*\mathbf{w} \subseteq \mathbf{L}$.

(2) If $\mathbf{L}$ is regular then there is a constant $\mathbf{c}$ such that for all $\mathbf{x} \in \mathbf{L}$ and for all sets $\mathbf{K}$ of $\mathbf{c}$ breakpoints on $\mathbf{x}$, there are breakpoints $\mathbf{i}, \mathbf{j}, \mathbf{h}, \mathbf{k} \in \mathbf{K}$ splitting $\mathbf{x}$ into $\mathbf{u}, \mathbf{v}, \hat{\mathbf{v}}, \tilde{\mathbf{v}}, \mathbf{w}$ with $\mathbf{u}\,\mathbf{v}^*\,\hat{\mathbf{v}}\,\tilde{\mathbf{v}}^*\,\mathbf{w} \subseteq \mathbf{L}$.

Property (2) can also be proven for block pumpable languages, property (1) is an open problem.

Theorem [Chak, Freivalds, Stephan and Tan 2016]
If $\mathbf{L}$ is block pumpable then there are $\mathbf{p}, \mathbf{p}' \geq \mathbf{2}$ such that for all $\mathbf{x}, \mathbf{y}$ with $\mathbf{xy} \in \mathbf{L}$ and all sets $\mathbf{I}$ of $\mathbf{p}$ breakpoints of $\mathbf{x}$ and $\mathbf{I}'$ of $\mathbf{p}'$ breakpoints of $\mathbf{y}$ there are pairs of breakpoints $\mathbf{i}, \mathbf{j} \in \mathbf{I}$ and $\mathbf{i}', \mathbf{j}' \in \mathbf{I}'$ such that $\mathbf{P}^*(\mathbf{x} : \mathbf{i}, \mathbf{j}) \cdot \mathbf{P}^*(\mathbf{y} : \mathbf{i}', \mathbf{j}') \subseteq \mathbf{L}$.

# Growth of Languages

A language $\mathbf{L}$ has polynomial growth iff there is a polynomial $\mathbf{p}$ such that, for all $\mathbf{n}$, $\mathbf{L}$ contains at most $\mathbf{p(n)}$ many words shorter than $\mathbf{n}$.

A language $\mathbf{L}$ has exponential growth iff there is are constants $\mathbf{a, b}$ such that $\mathbf{L}$ contains, for every $\mathbf{n}$, at least $\mathbf{2^n}$ words shorter than $\mathbf{a \cdot n + b}$.

Theorem [Bridson and Gilman 2002, Incitti 2001]
Every regular language has either polynomial or exponential growth.
Every context-free language has either polynomial or exponential growth.

Examples
Polynomial growth: $\{\mathbf{0}\}^*$, $\{\mathbf{00}\}^* \cdot \{\mathbf{111}\}^*$.
Exponential growth: $\{\mathbf{0, 1}\}^*$, $\{\mathbf{000, 111}\} \cdot \{\mathbf{00, 11}\}^*$.

# Results

**Theorem**
If $\mathbf{L}$ is block pumpable and $\mathbf{H}$ is a regular language of polynomial growth then $\mathbf{L} \cap \mathbf{H}$ is a regular language of polynomial growth.

**Corollary**
A block pumpable language of polynomial growth is regular.

**Theorem**
If $\mathbf{L}$ is block pumpable and $\mathbf{H}$ is a context-free language of polynomial growth then $\mathbf{L} \cap \mathbf{H}$ is a context-free language of polynomial growth.

**Proof.** A polynomial language of polynomial growth is a subset of a regular language of polynomial growth, the latter intersected with $\mathbf{L}$ is regular and that regular set intersected with $\mathbf{H}$ is $\mathbf{L} \cap \mathbf{H}$ and is context-free. As a subset of a set with polynomial growth, it has polynomial growth.

# Blockpumpable Structures

A structure $(\mathbf{A}, \mathbf{R_1}, \ldots, \mathbf{R_n})$ is automatic iff the domain $\mathbf{A}$ and each relation $\mathbf{R_m}$ represented as the set $\{\mathbf{conv}(\mathbf{x_1}, \ldots, \mathbf{x_k}) : (\mathbf{x_1}, \ldots, \mathbf{x_k}) \in \mathbf{R_m}\}$ is regular.

Here the convolution is obtained by aligning of each of $\mathbf{x_1}, \ldots, \mathbf{x_k}$ the symbols at the same position to one new symbol, if some of these do not exist (due to the input being shorter), the corresponding component is replaced by the special symbol $\#$. So

$$\mathbf{conv}(\mathbf{001}, \mathbf{12211}) = \binom{\mathbf{0}}{\mathbf{1}}\binom{\mathbf{0}}{\mathbf{2}}\binom{\mathbf{1}}{\mathbf{2}}\binom{\#}{\mathbf{1}}\binom{\#}{\mathbf{1}}$$

A structure $(\mathbf{A}, \mathbf{R_1}, \ldots, \mathbf{R_n})$ is effectively block pumpable iff the domain and the convolution of each relation is a block pumpable recursive set.

# Examples

## Integers

The integers with the relations $<$ given by the ordering and $\{(\mathbf{x},\mathbf{y},\mathbf{z}) : \mathbf{x}+\mathbf{y}=\mathbf{z}\}$ form an automatic structure.

## Rationals

The rationals with addition do not form an automatic structure.

## Block pumpable domain

The set $\mathbf{L_{exp}} \cup \mathbf{L_{sc}}$ together with the lexicographic order and predicates $\mathbf{P_a}$ telling whether a string ends with $\mathbf{a}$ is an effectively block pumpable structure.

## Block pumpable relation

Let $\mathbf{R}(\mathbf{x},\mathbf{y})$ be true iff $\mathbf{x}$ contains a square or $\mathbf{y}$ contains a square or $\mathbf{2} \cdot |\mathbf{x}| \neq |\mathbf{y}|$. The structure $(\{\mathbf{0},\mathbf{1},\mathbf{2}\}^*, \mathbf{R})$ is effectively block pumpable.

# Decidability

Theorem [Chak, Freivalds, Stephan and Tan 2016]
If $(\mathbf{A}, \mathbf{R_1}, \ldots, \mathbf{R_n})$ is an effective block pumpable structure and $\Phi$ a predicate which is an existentially quantified disjunction of conjunctions of relations in the structure then $\mathbf{\Phi}$ is decidable.

Idea: Witnesses of existentially quantified variables in a formula $\mathbf{\Phi}$ can be assumed to have a certain maximum length which depends on the pumping constants of the various relations involved as well as on the pumping constants of derived relations which are obtained by considering conjunctions or disjunctions of other block pumpable relations.

The decidability of the relations and domain can then be used to test out one by one the various possibility below the corresponding length bound.

# Positive Theory

The positive first-order theory of a structure consists of quantified formulas which do not use the negation. This theory can be undecidable.

Example [Chak, Freivalds, Stephan and Tan 2016]
Let $\mathbf{A} = \{\mathbf{x3y} : \mathbf{x}, \mathbf{y} \in \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}^* : \mathbf{x}$ contains a square or $\mathbf{y}$ contains a square or $|\mathbf{x}|$ is different from $\mathbf{k}_{|\mathbf{y}|}\}$ where $\mathbf{k_0}, \mathbf{k_1}, \ldots$ is a recursive one-one enumeration of the halting problem $\mathbf{K}$.
$\mathbf{R_1}(\mathbf{v}, \mathbf{w})$ is true iff $\mathbf{v} \in \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}^* \cdot \{\mathbf{3}\}$ and $\mathbf{w}$ prop. extends $\mathbf{v}$.
$\mathbf{R_2}(\mathbf{v}, \mathbf{w})$ is true iff $\mathbf{v0} \neq \mathbf{w}$.
Let $\mathbf{\Phi}(\mathbf{x}) = \exists \mathbf{v} \forall \mathbf{w}\, [\mathbf{R_1}(\mathbf{a}, \mathbf{v}) \wedge \mathbf{R_2}(\mathbf{v}, \mathbf{w})]$.
Now $\mathbf{n} \in \mathbf{K} - \{\mathbf{k_0}\}$ iff $\mathbf{\Phi}(\mathbf{a})$ holds for a square-free $\mathbf{a} \in \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}^{\mathbf{n}} \cdot \{\mathbf{3}\}$.
It is impossible to decide whether $\mathbf{\Phi}(\mathbf{a})$ is true.

# Concrete Bounds

The language of non-squares is context-free and the language of irreducible words $\mathbf{x}$ ($\mathbf{x} \neq \mathbf{y}^{\mathbf{n}}$ for all $\mathbf{n} > \mathbf{1}$) is perhaps context-free; this is a famous open problem.

These two languages are not positively block pumpable: for sufficiently large $\mathbf{m}, \mathbf{n}$ with $\mathbf{n} - \mathbf{m}$ being a sufficiently large factorial, one can pump the word $\mathbf{0^m 1 0^n 1}$ to get the square $\mathbf{0^n 1 0^n 1}$. However, these languages are block cancellable.

Theorem [Sung 2016]
The language of non-squares is block cancellable with constant $\mathbf{7}$ or less.
The language of irreducible words is block cancellable with constant $\mathbf{13}$ or less.

# Bounds for Non-Regularity

Pumping constants for block pumpable, block cancellable and positively block pumpable languages are at least $\mathbf{2}$, as one can otherwise not define a pump at all.

Example: The language $\mathbf{L_{sc}}$ of square-containing words is, for alphabet size at least $\mathbf{3}$, positively block pumpable with constant $\mathbf{2}$ and not regular.

Fact: If $\mathbf{L}$ is block cancellable with constant $\mathbf{2}$ then it contains with each word all subsequence words of it; it is known that such a language is regular.

Theorem [Sung 2016]: A language $\mathbf{L}$ is $\mathbf{L}$ is block cancellable with constant $\mathbf{2}$ iff $\mathbf{L}$ is the union of some sets of the form $\mathbf{A}^*$ with all words in $\mathbf{A}$ having length $\mathbf{1}$.
Let $\mathbf{H} = \{\mathbf{3v3w3} : \mathbf{v}, \mathbf{w} \in \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}^*, \mathbf{v} \neq \mathbf{w}, |\mathbf{v}| = |\mathbf{w}|$ and $\mathbf{v}, \mathbf{w}$ are square-free$\}$ and $\mathbf{L} = \{\mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}\}^* - \mathbf{H}$. $\mathbf{L}$ is block pumpable with constant $\mathbf{3}$ and not regular.

# Open Problems

1. Many bounds on the block pumping constant are computed using Ramsey's Theorem of Pairs, for example when forming intersections of other block pumpable languages. This gives very bad bounds, which do not come up when dealing with regular languages and the bounds stemming from the size of automata. Therefore one might ask whether there are better bounds.

2. If $L$ is block pumpable, can one select in sufficiently large sets of breakpoints two neighbouring pumps which can be pumped independently?

3. Are there block cancellable languages which have neither polynomial nor exponential growth?

4. If $f : A \to B$ is a block pumpable function from a block pumpable set $A$ to a block pumpable set $B$, is there an automatic function which coincides with $f$ on $A$?

# Summary

Block pumpable languages $L$ are a generalisation of regular languages with respect to one-sided applications of the block pumping lemma (only to members of $L$).

Block pumpable languages are closed under union, concatenation, intersection and the image of transductions; but not closed under complement, Kleene star.

Block pumpable languages allow multiple pumping.

Block pumpable languages of polynomial growth are regular. The intersection of a block pumpable language and a context-free language of polynomial growth is context-free.

One can define block pumpable structures using an effective version of block pumpable languages and relations and have some weak decidability properties.