Regular languages over countable words

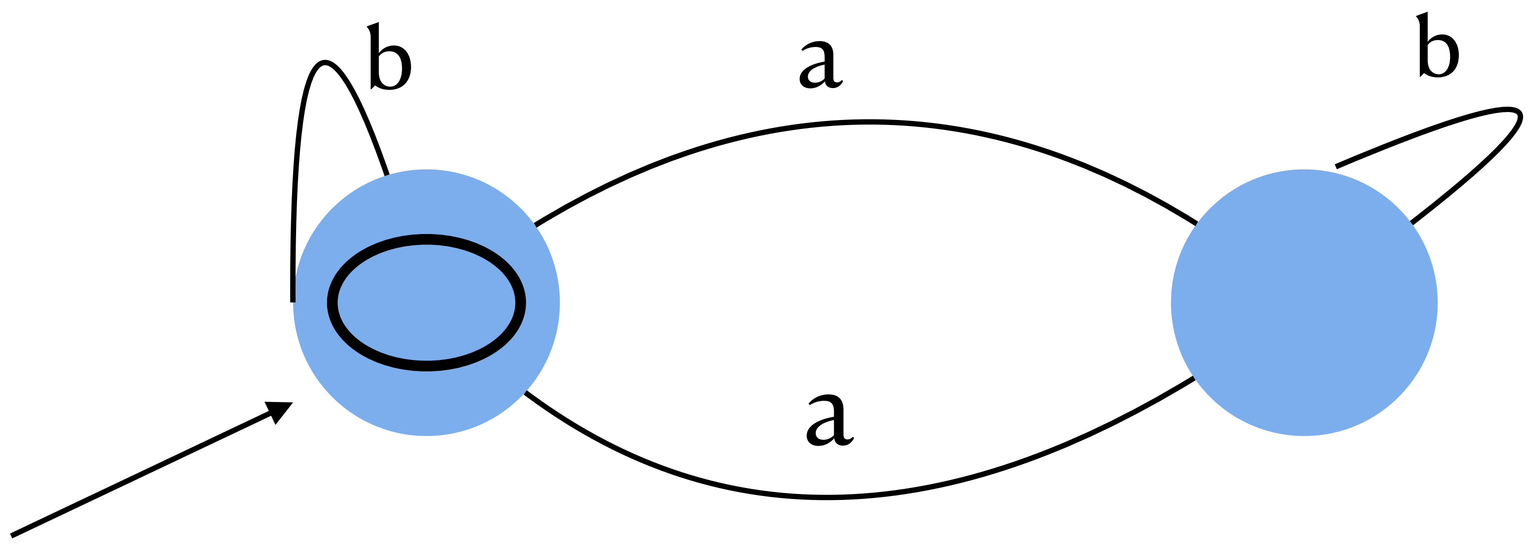
Formal methods update meeting 2021

Sreejith A. V. (IIT Goa)

Regular languages

Descriptive formalism vs Computational models

- Consider the language: even number of a's
- Regular expression: b*(a b* a b*)*
- Monadic second order logic: $\exists X\ \mathit{first}_a(X) \wedge \neg \mathit{last}_a(X) \wedge$
  "if y in X then a(y) and next a-position not in X"
- Effective translation of logic to automata is central to many verification tools (eg. SPIN)
- [Vardi, Wolper] Further refined and extended to practical specification languages like LTL.
- [Buchi, Rabin] Extended to omega words



Deterministic Finite Automata (DFA)

⇕ [Kleene]

Regular expressions

⇕ [BET]

Monadic second order logic

Regular languages and Algebraic structures

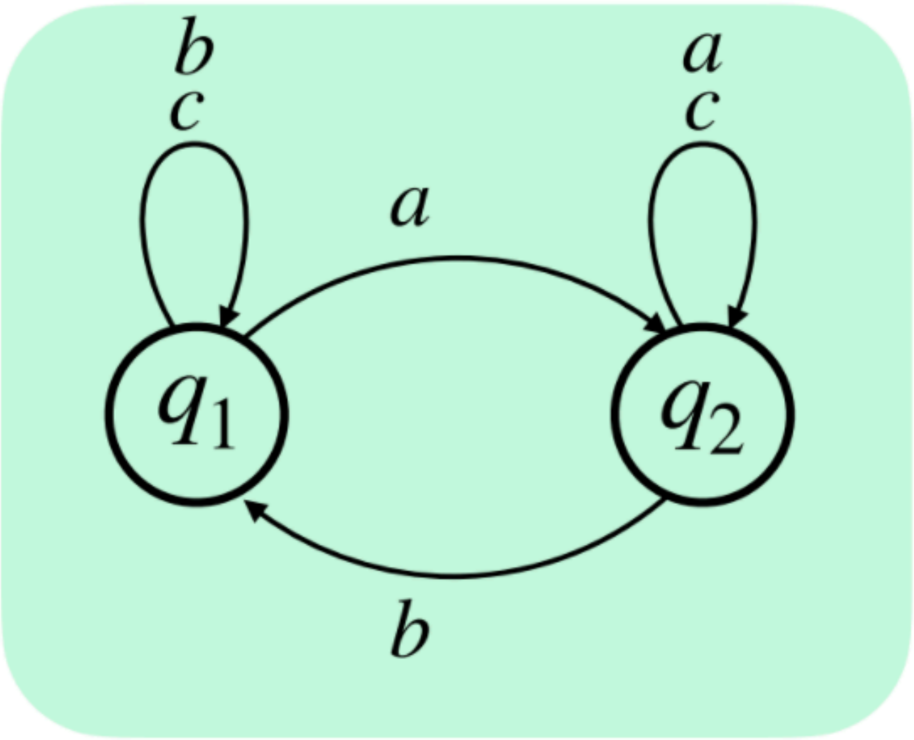
A refined understanding of regular languages

- [Rabin-Scott, Myhill-Nerode] Regular languages ⇔ recognised by *finite monoid*
- Monoids are “close cousins” of finite automata: a set with an associative operation
- [Schutzenberger, McNaughton-Papert, Kamp]
  Star free Exp ⇔ counter free automata ⇔ aperiodic monoids ⇔ FO-logic ⇔ LTL
- Algebraic characterisations of many other logical fragments known.
- Leads to decision procedures for checking definability in a logical fragment/formalism.

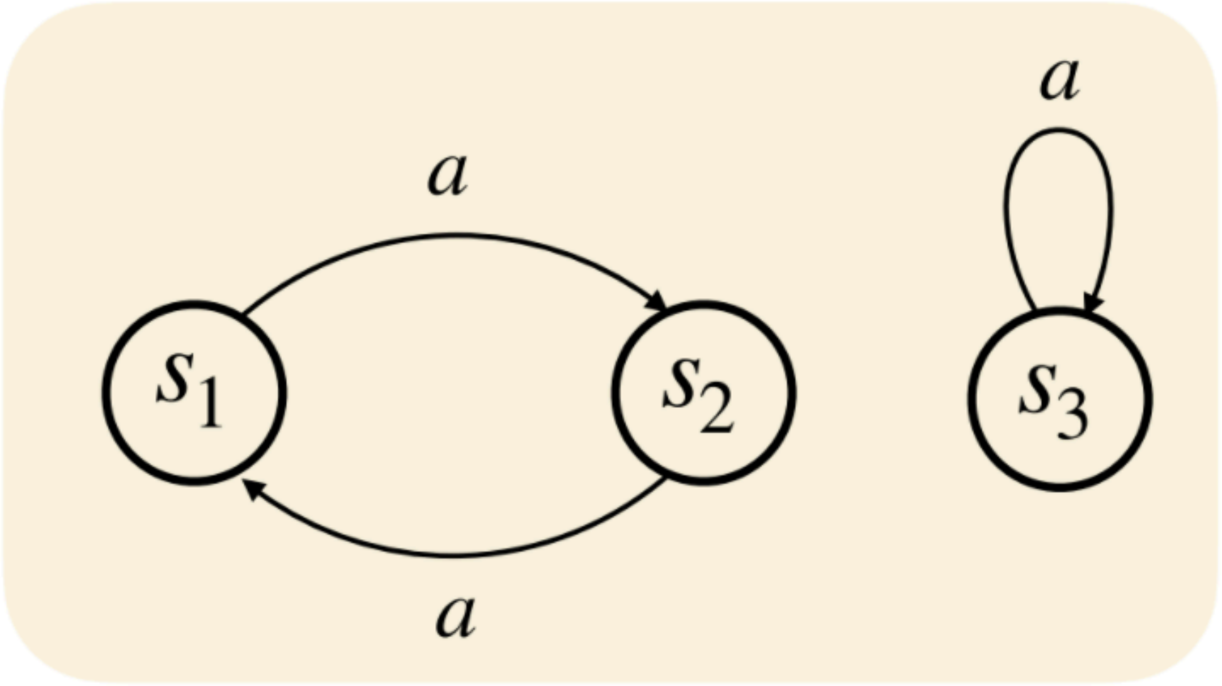
Decomposition of automata/monoids

Several simple machines can be used to build complex machines

- [Krohn-Rhodes] Any automata can be decomposed into a *cascade product* of *simpler automata*.
- Any automata can be build by connecting *simple automata* in *parallel or series*.
- Any monoid *divides* a *wreath product* of *simpler monoids*.
- *Aperiodic monoid* ⇔ monoid is a *wreath product* of a unique monoid U_2.
- The decomposition theorem allows inductive reasoning of automata/monoids.



Two state reset automata U_2



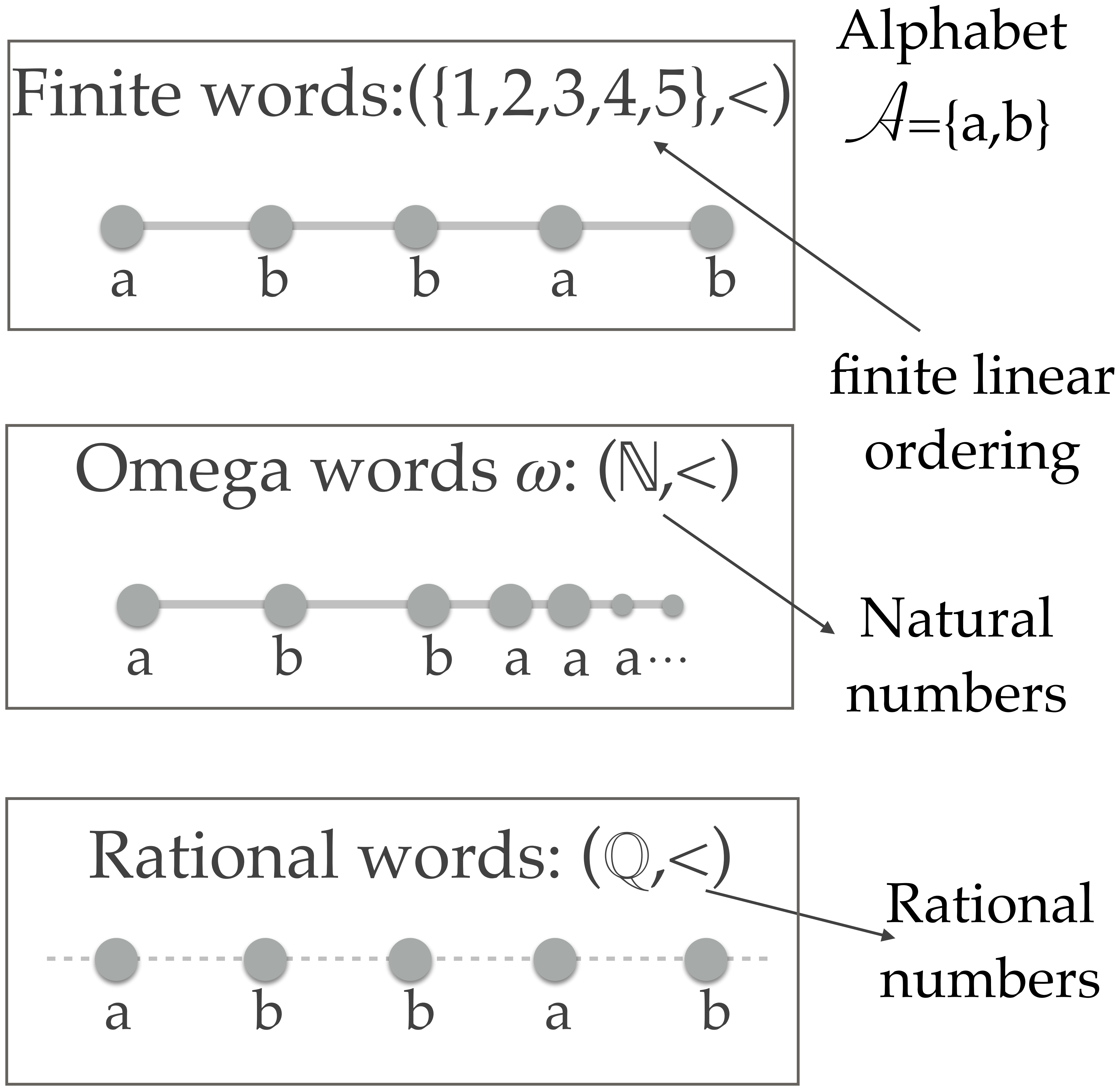
Permutation automata G

Countable words

A generalisation of finite and omega words

- Our aim is to extend the study of regular languages to *countable words*.
- (L, <) is a *countable linear ordering* if L is a countable set and < is a total order on L.
- A *countable word* over alphabet A is a map w: L ->A
- An *MSO definable language of countable words* is called regular language.

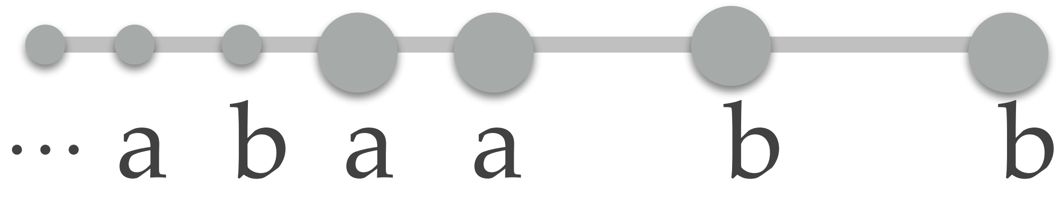
eg 1. letter b occur densely: $\forall x, y \; \exists z \; (x < z < y) \wedge b(z)$
eg 2. infinitely many b: $\forall X \; (\mathit{finite}(X) \Rightarrow \exists z \notin X \wedge b(z))$



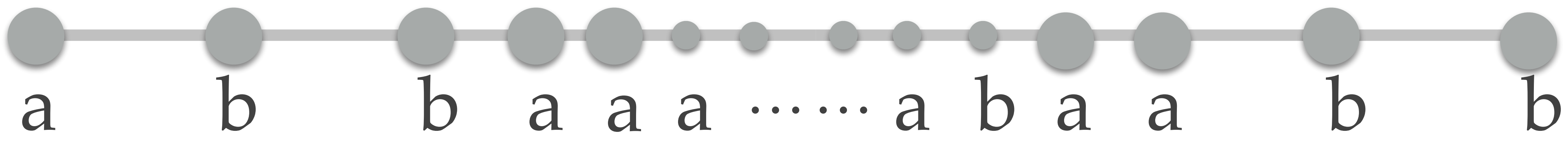
Examples of countable words

(recall: countable words are mappings from countable orderings to a finite alphabet)

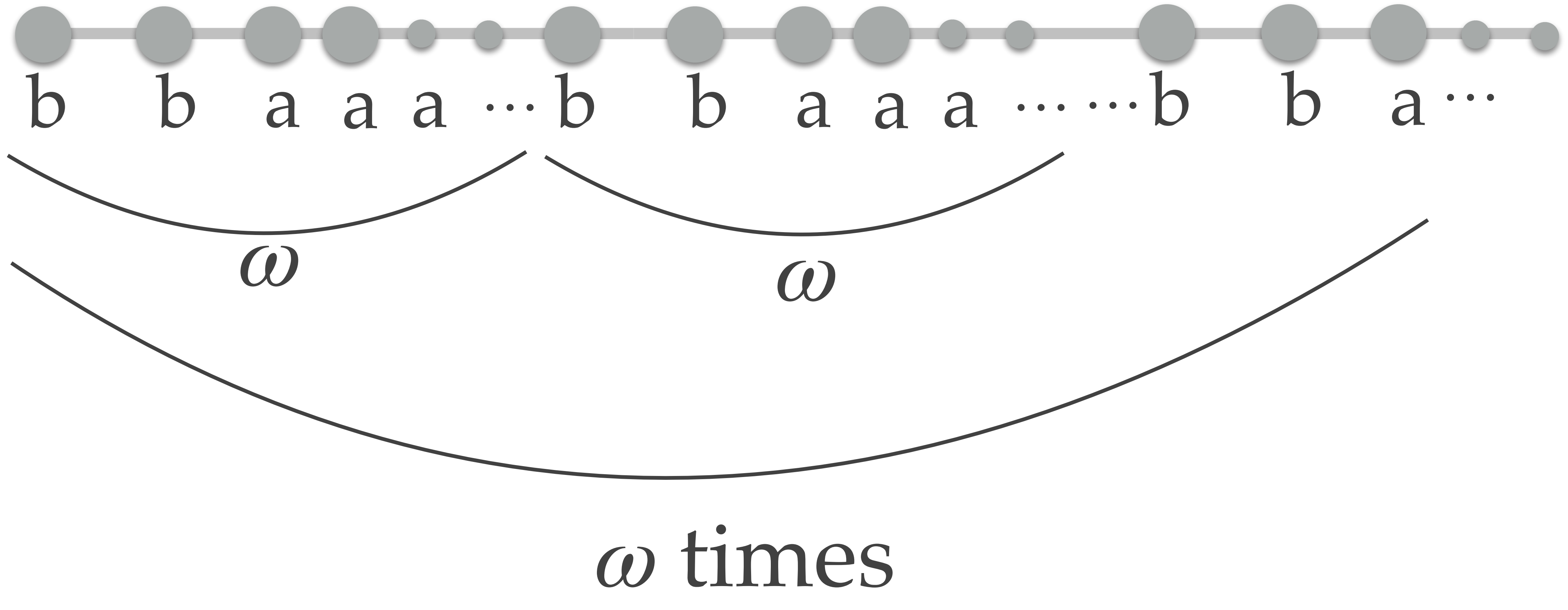
Omega* words ω^*: $(-\mathbb{N},<)$



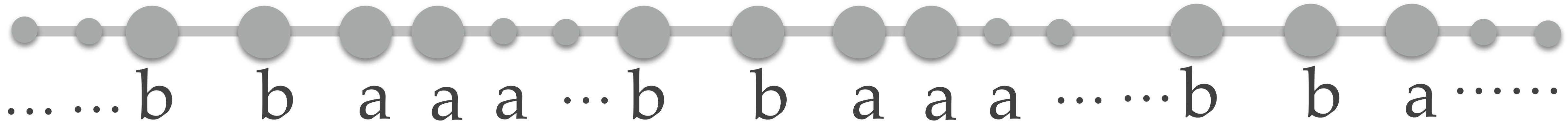
Domain $\omega+\omega^*$



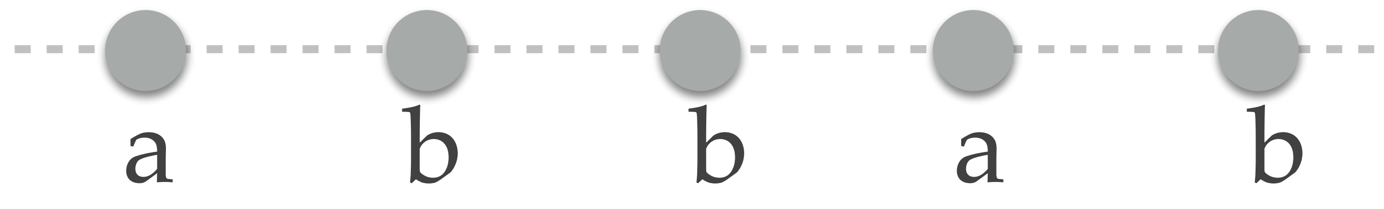
well ordered words (ordinals)



scattered ordering: No *dense sub-ordering*



Dense ordering: Example, Rationals $(\mathbb{Q},<)$



Monadic second order logic

and fragments of MSO

- Order relation: $x < y$, $x = y$, $x > y$
- Letter Predicates: $a(x)$, $b(x)$, where a and b are letters
- Boolean closure (and, or, negation): $\phi_1 \lor \phi_2, \phi_1 \lor \phi_2, \neg\phi$
- First-order quantification: $\exists\, x\, \phi\,(x)$

First-order logic

- Monadic second order quantification: $\exists\, X\, \phi(X)$
- Set membership: $x \in X$
- Based on type of set quantifiers X, other natural logics definable
- Weak MSO (WMSO): quantify only over finite sets
- FO[cut]: quantify only over cuts

Example Regular languages

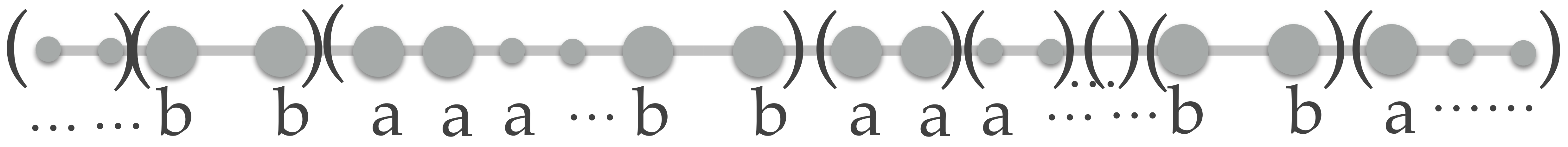
MSO defines languages of countable words

- set of all dense words? First order - for all x, y there exists a z such that $x < z < y$
- ω sequence of a's to the right of the word? FO - $\forall x$ "if a(x) then there $\exists\, y > x \wedge a(y)$"
- infinitely many a's? WMSO - for all finite set X, there is an x (not in X) such that a(x)
- sub ordering with ω^* many a's? FO[cut] definable
- there is no dense set of a's?
- Many interesting questions -
  - Is language L first order definable?
  - Is WMSO strictly less powerful than MSO?
  - Is there a characterisation for each of these logics?

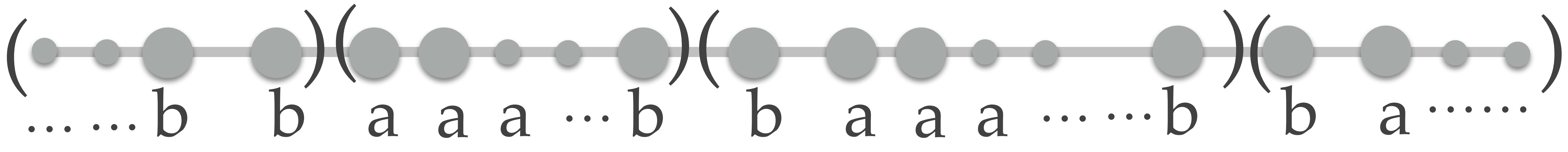
An algebra for countable words

monoids satisfying generalised associativity

- A *o-monoid* (M,π) is a set M equipped with a product π: countable words(M) -> M that satisfies *generalised associativity*.
- Associativity: (a.(b.c)) = ((a.b).c), i.e $\pi(a\ \pi(bc)) = \pi(\pi(ab)c)$
- *Generalized associativity* (informally): Whichever way you bracket an infinite word the product should be the same.

1. 

Both should compute the same product.

2. 

Recognizability through algebra

MSO definability ⇔ recognisability by o-monoids

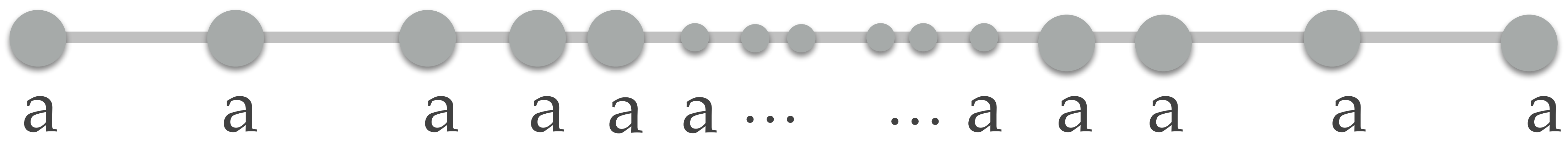
- [Shelah; Carton-Colcombet-Puppis] A set of countable words is definable in MSO if and only if it is recognisable by a *finite o-monoid.*
- Corollary [Shelah] Emptiness (or universality) of an MSO formula is decidable.
- Corollary MSO hierarchy collapse to the second level.

First order logic vs Gaps

(Finiteness is not FO definable)

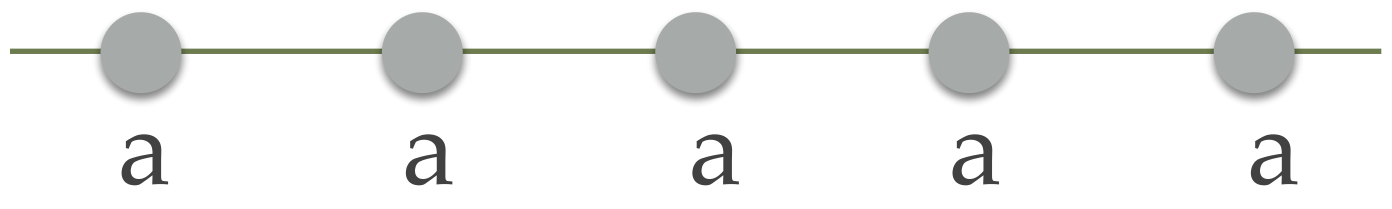
- [Schützenberger] A set of finite words is definable in FO if and only if it is recognised by an aperiodic monoid.
- [Bès & Carton] A set of countable words is definable in FO only if every idempotent e satisfies equation $e^{\omega} \circ e^{\omega^*} = e$.

$a^{\omega} a^{\omega^*}$



is FO equivalent to

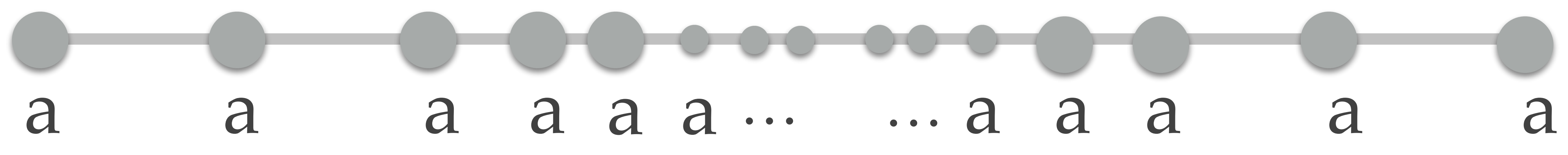
long finite sequence of a's



First order logic vs Aperiodicity

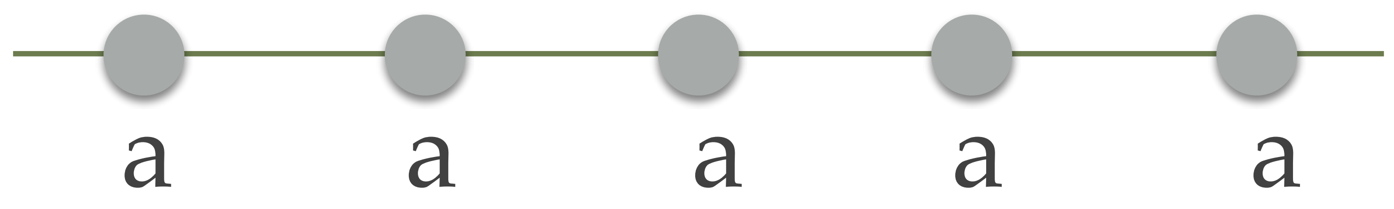
(aperiodicity does not imply first-order definability)

$a^{\omega}a^{\omega^*}$



is FO equivalent to

long finite sequence of a's



- Corollary. FO cannot detect gaps.
  FO cannot recognise the set of all finite words.
- Corollary. Implies aperiodicity (or group freeness), i.e. $\exists$ n, $x^n = x^{n+1}$
  Proof. Let $a^{\omega}a^{\omega^*} = a^n$. Then $a^{n+1} = a \circ a^n = a \circ a^{\omega}a^{\omega^*} = a^{\omega}a^{\omega^*} = a^n$.
- Remark. The above equation is not sufficient to capture FO

Effective characterisation for fragments of MSO

o-monoids give a refined understanding of languages

- [Colcombet, S. ICALP15] There is an effective characterisation of FO/WMSO/FOCUT/WMSOCUT/FOSCATTERED definable countable languages.
- Corollary. Can answer questions like this.
  - Is WMSO a strict subset of MSO?
  - Is a language FO or WMSOCUT definable?
- [Manuel, S. MFCS17] An effective algebraic characterisation of FO2.
- Corollary. If an FO2 formula has a satisfying model, then it has a satisfying model which is a scattered linear ordering.

Decomposition of o-monoids

(Krohn-Rhodes style)

- [Adsul, Sarkar, S. LICS19, FCT21] Decomposition theorem for o-monoids equivalent to various fragments of MSO
  - First-order logic
  - Two variable first-order logic
  - linear temporal logic
  - some natural logics more expressive than first-order logic
- Corollary. Regular expression for first-order logic
- Corollary. Aperiodic o-monoids cannot be decomposed using finite number of simpler o-monoids (different from the case of monoids)

Summary and Research directions

Many questions remain unanswered

- Regular languages over countable words have rich characterisations.
- There are many unanswered questions
  - Decomposition theorem for the full o-monoid (corresponds to MSO).
  - Proper study on regular expressions.
  - Satisfiability/model checking algorithms.
  - Applications to verification.
- Other research directions
  - Going beyond countable words.
  - Shelah's Open question. Over reals MSO quantification over Borel sets is decidable.

Thank You