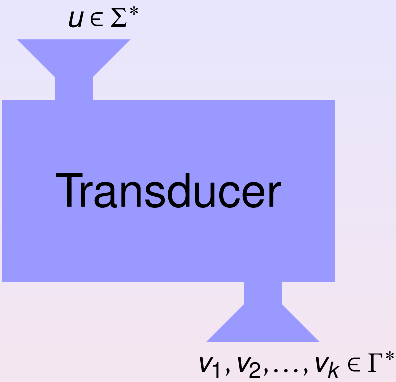
Characterizing the Valuedness of Two-way Finite Transducers

Di-De Yen and Hsu-Chun Yen

Department of Electrical Engineering
National Taiwan University

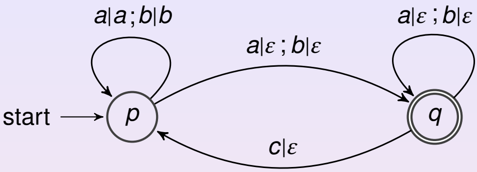
Outline

1. Introduction and definitions
2. The valuedness problem of two-way finite transducers
3. A sufficient and necessary condition for the valuedness problem
4. Future work



- The Finite-valuedness Problem:
  Is there a constant C such that $k \leq C$ for all u?
- Main Result:
  We give a necessary and sufficient condition for the finite-valuedness of two-way finite transducers!

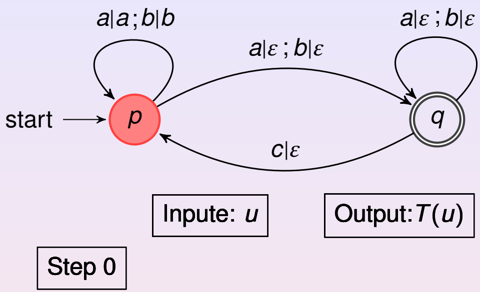
One-way Finite Transducer (1FT)



1FT $T = (Q, \Sigma, \Gamma, q_0, F, \Delta)$

- Q: a finite set of states
  $\Rightarrow Q = \{p, q\}$
- Σ: a finite input alphabet
  $\Rightarrow \Sigma = \{a, b, c\}$
- Γ: a finite output alphabet
  $\Rightarrow \Gamma = \{a, b\}$
- q_0: an initial state
  $\Rightarrow q_0 = p$
- F: a set of final states
  $\Rightarrow F = \{q\}$
- Δ: a finite set of transitions
  $(\Delta \subseteq Q \times \Sigma \times Q \times \Gamma^*)$

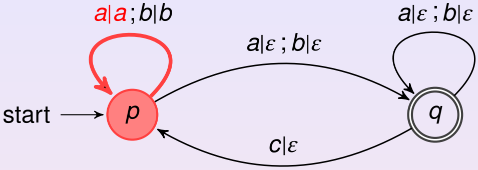
One-way Finite Transducer (1FT)



1FT $T = (Q, \Sigma, \Gamma, q_0, F, \Delta)$

- Q: a finite set of states
  $\Rightarrow Q = \{p, q\}$
- Σ: a finite input alphabet
  $\Rightarrow \Sigma = \{a, b, c\}$
- Γ: a finite output alphabet
  $\Rightarrow \Gamma = \{a, b\}$
- q_0: an initial state
  $\Rightarrow q_0 = p$
- F: a set of final states
  $\Rightarrow F = \{q\}$
- Δ: a finite set of transitions
  $(\Delta \subseteq Q \times \Sigma \times Q \times \Gamma^*)$

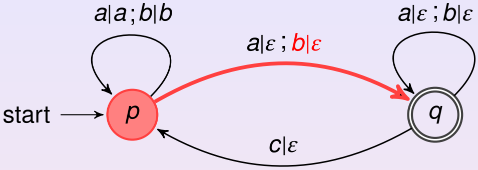
One-way Finite Transducer (1FT)



Step 1	Inpute: u	Output: $T(u)$
	a	a

1FT $T = (Q, \Sigma, \Gamma, q_0, F, \Delta)$

- Q: a finite set of states
  $\Rightarrow Q = \{p, q\}$
- Σ: a finite input alphabet
  $\Rightarrow \Sigma = \{a, b, c\}$
- Γ: a finite output alphabet
  $\Rightarrow \Gamma = \{a, b\}$
- q_0: an initial state
  $\Rightarrow q_0 = p$
- F: a set of final states
  $\Rightarrow F = \{q\}$
- Δ: a finite set of transitions
  $(\Delta \subseteq Q \times \Sigma \times Q \times \Gamma^*)$

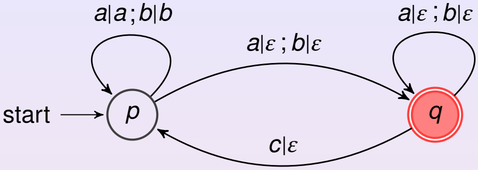


Step 2	Inpute: u	Output: $T(u)$
	ab	aε

1FT $T = (Q, \Sigma, \Gamma, q_0, F, \Delta)$

- Q: a finite set of states $\Rightarrow Q = \{p, q\}$
- Σ: a finite input alphabet $\Rightarrow \Sigma = \{a, b, c\}$
- Γ: a finite output alphabet $\Rightarrow \Gamma = \{a, b\}$
- q_0: an initial state $\Rightarrow q_0 = p$
- F: a set of final states $\Rightarrow F = \{q\}$
- Δ: a finite set of transitions $(\Delta \subseteq Q \times \Sigma \times Q \times \Gamma^*)$

One-way Finite Transducer (1FT)

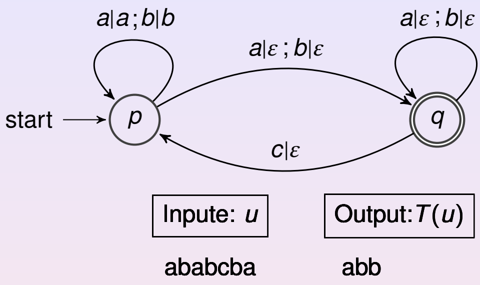


Step 7	Inpute: u	Output: $T(u)$
	ababcba	abb

1FT $T = (Q, \Sigma, \Gamma, q_0, F, \Delta)$

- Q: a finite set of states
  $\Rightarrow Q = \{p, q\}$
- Σ: a finite input alphabet
  $\Rightarrow \Sigma = \{a, b, c\}$
- Γ: a finite output alphabet
  $\Rightarrow \Gamma = \{a, b\}$
- q_0: an initial state
  $\Rightarrow q_0 = p$
- F: a set of final states
  $\Rightarrow F = \{q\}$
- Δ: a finite set of transitions
  $(\Delta \subseteq Q \times \Sigma \times Q \times \Gamma^*)$

One-way Finite Transducer (1FT)

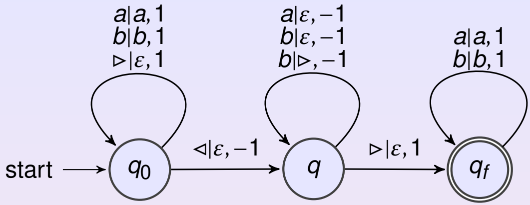


1FT $T = (Q, \Sigma, \Gamma, q_0, F, \Delta)$

- Q: a finite set of states
  $\Rightarrow Q = \{p, q\}$
- Σ: a finite input alphabet
  $\Rightarrow \Sigma = \{a, b, c\}$
- Γ: a finite output alphabet
  $\Rightarrow \Gamma = \{a, b\}$
- q_0: an initial state
  $\Rightarrow q_0 = p$
- F: a set of final states
  $\Rightarrow F = \{q\}$
- Δ: a finite set of transitions
  $(\Delta \subseteq Q \times \Sigma \times Q \times \Gamma^*)$

- $T(u)$: the set of corresponding outputs of u
  $\Rightarrow abb \in T(ababcba)$
- $R(T)$: $\{(u, v) | v \in T(u), u \in \Sigma^*\}$
  $\Rightarrow R(T) = \{(u_1 c u_2, prefix(u_1) prefix(u_2)) | u_1, u_2 \in a, b^*\}$

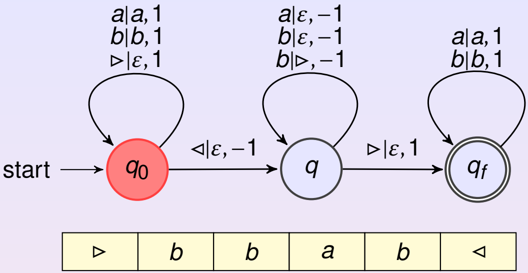
Two-way Finite Transducer (2FT)



2FT $T = (Q, \Sigma, \Gamma, q_0, F, \Delta)$

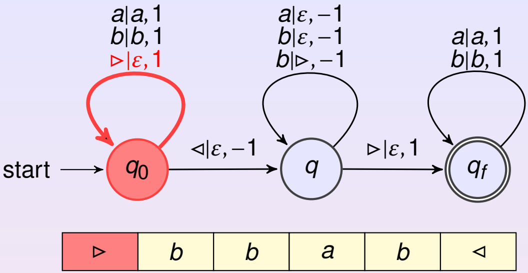
- $\Delta \subseteq Q \times (\Sigma \cup \{\rhd, \lhd\}) \times Q \times \Gamma^* \times \{-1, 1\}$

Two-way Finite Transducer (2FT)



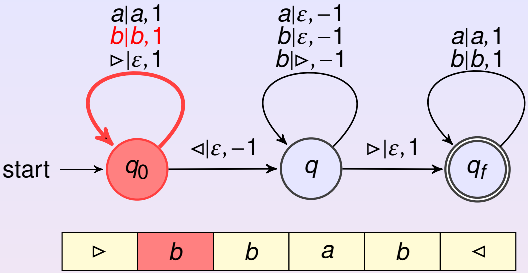
π: q_0

Two-way Finite Transducer (2FT)



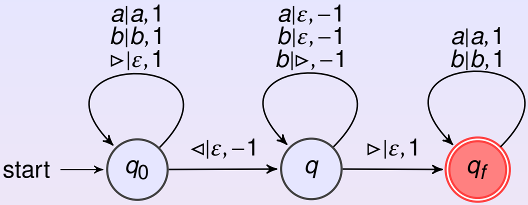
$\pi: q_0 \xrightarrow{\varepsilon} q_0$

Two-way Finite Transducer (2FT)



$$\pi : q_0 \xrightarrow{\varepsilon} q_0 \xrightarrow{b} q_0$$

Two-way Finite Transducer (2FT)



▷	*b*	*b*	*a*	*b*	◁

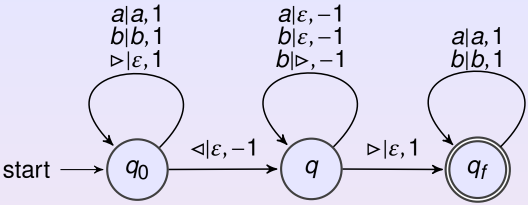
$$\pi: q_0 \xrightarrow{\varepsilon} q_0 \xrightarrow{b} q_0 \xrightarrow{b} q_0 \xrightarrow{a} q_0 \xrightarrow{b} q_0 \xrightarrow{\varepsilon}$$

$$q \xleftarrow{\varepsilon} q \xleftarrow{\varepsilon} q \xleftarrow{\varepsilon} q \xleftarrow{\varepsilon} q$$

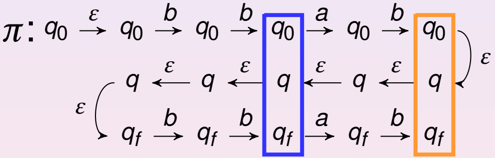
$$\xrightarrow{\varepsilon} q_f \xrightarrow{b} q_f \xrightarrow{b} q_f \xrightarrow{a} q_f \xrightarrow{b} q_f$$

- $bbabbbab \in T(bbab)$.
- $R(T) = \{(u, uu) | u \in a, b^*\}$.

Two-way Finite Transducer (2FT)

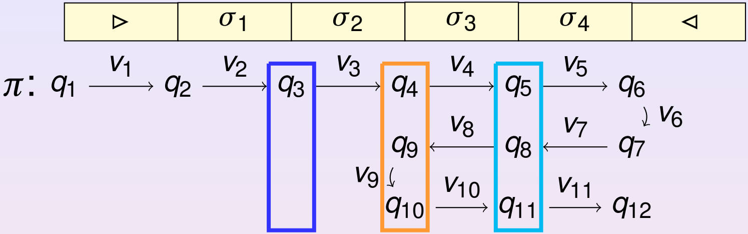


$\triangleright$	b	b	a	b	$\triangleleft$



- $bbabbbab \in T(bbab)$.
- $R(T) = \{(u,uu) | u \in a,b^*\}$.
- Crossing sequence: a sequence in $(Q \times \{-1,1\})^*$.
- $C_\pi[3] = C_\pi[6] = (q_0,1)(q,-1)(q_f,1)$.

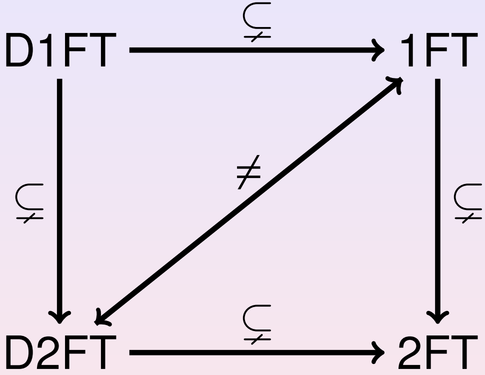
Another Example



Example

- $C_\pi[3] = (q_3, 1)$.
- $C_\pi[4] = (q_4, 1)(q_9, 1)(q_{10}, 1)$.
- $C_\pi[5] = (q_5, 1)(q_8, -1)(q_{11}, 1)$.

Expressive Power



Notions of Valuedness

1. Single-valued: $|T(u)| \leq 1, \; \forall u \in \Sigma^*$.
2. k-valued: $|T(u)| \leq k, \; \forall u \in \Sigma^*$.
3. Finite-valued: k-valued, for some $k \in N$.
4. Infinite-valued: not finite-valued.

Notions of Valuedness

1. Single-valued: $|T(u)| \leq 1,\ \forall u \in \Sigma^*$.
2. k-valued: $|T(u)| \leq k,\ \forall u \in \Sigma^*$.
3. Finite-valued: k-valued, for some $k \in N$.
4. Infinite-valued: not finite-valued.

	1FT	2FT
Single-valued	D [Blattner, et al.,77]	D [Culik, et al.,87]
k-valued	D [Gurari, et al.,83]	D [Culik, et al.,86]
Finite-valued	D [Weber,90]	?

Notions of Valuedness

1. Single-valued: $|T(u)| \leq 1, \ \forall u \in \Sigma^*$.
2. k-valued: $|T(u)| \leq k, \ \forall u \in \Sigma^*$.
3. Finite-valued: k-valued, for some $k \in N$.
4. Infinite-valued: not finite-valued.

	1FT	2FT
Single-valued	D [Blattner, et al.,77]	D [Culik, et al.,87]
k-valued	D [Gurari, et al.,83]	D [Culik, et al.,86]
Finite-valued	D [Weber,90]	?

We give a 'simple' necessary and sufficient condition for the infinite-valuedness of two-way finite transducers!

Characteristics of Infinite-valuedness

Theorem (Weber, 90)

A 1FT T is infinite-valued iff it satisfies criterion IV1 or IV2.

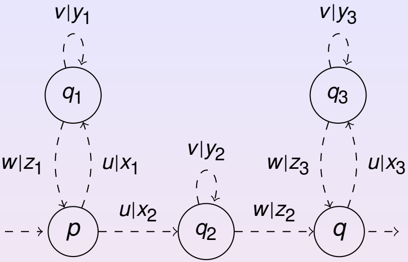
- Finite Automata:

  (Crossing Sequence, Two-way)

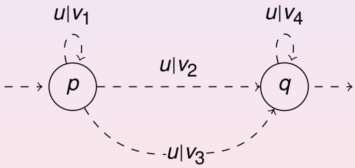
  $\Updownarrow$

  (State, One-way)

- Crossing sequence version of Weber's criteria ?!



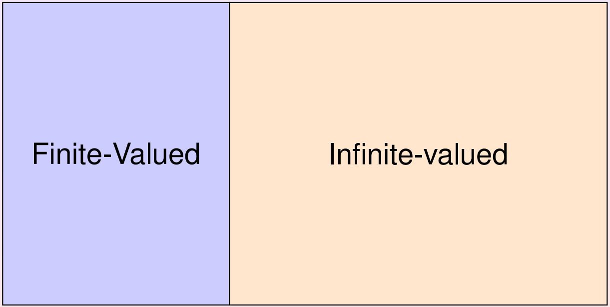
(a) IV1: $|y_1| \neq |y_2|$



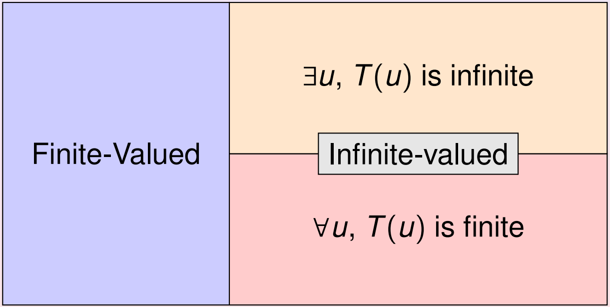
(b) IV2: $v_1 \neq \epsilon$ and $\exists i, v_2(i) \neq v_3(i)$

- In general, there are infinitely many crossing sequences!

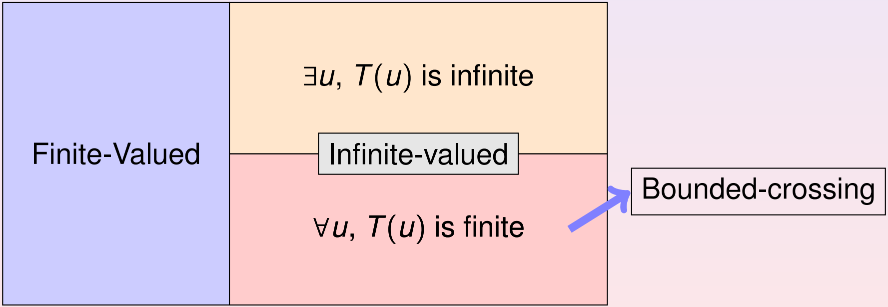
- In general, there are infinitely many crossing sequences!



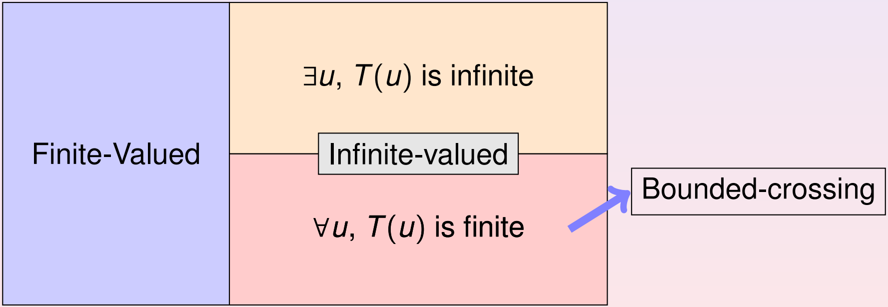
- In general, there are infinitely many crossing sequences!



- In general, there are infinitely many crossing sequences!
- $\exists u$, $T(u)$ is infinite iff there exists a π on u with a loop along which the output is not empty.



- In general, there are infinitely many crossing sequences!
- $\exists u$, $T(u)$ is infinite iff there exists a π on u with a loop along which the output is not empty.
- From now on, we assume all 2FTs are bounded-crossing.



Transitivity

- Known Results:
  1FT: $p \xrightarrow{u_1|v_1} q$ and $q \xrightarrow{u_2|v_2} r \Rightarrow p \xrightarrow{u_1 u_2|v_1 v_2} r$.
  2FA: $\mathbf{c} \xrightarrow{u_1} \mathbf{d}$ and $\mathbf{d} \xrightarrow{u_2} \mathbf{e} \Rightarrow \mathbf{c} \xrightarrow{u_1 u_2} \mathbf{e}$

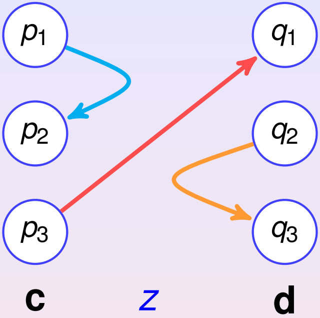
Transitivity

- Known Results:
  1FT: $p \xrightarrow{u_1|v_1} q$ and $q \xrightarrow{u_2|v_2} r \Rightarrow p \xrightarrow{u_1 u_2|v_1 v_2} r$.
  2FA: $\mathbf{c} \xrightarrow{u_1} \mathbf{d}$ and $\mathbf{d} \xrightarrow{u_2} \mathbf{e} \Rightarrow \mathbf{c} \xrightarrow{u_1 u_2} \mathbf{e}$
- Can we derive a "similar result" for *2FT*s?

Transitivity

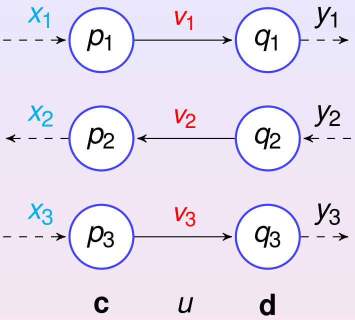
- Known Results:
  1FT: $p \xrightarrow{u_1|v_1} q$ and $q \xrightarrow{u_2|v_2} r \Rightarrow p \xrightarrow{u_1 u_2|v_1 v_2} r$.
  2FA: $\mathbf{c} \xrightarrow{u_1} \mathbf{d}$ and $\mathbf{d} \xrightarrow{u_2} \mathbf{e} \Rightarrow \mathbf{c} \xrightarrow{u_1 u_2} \mathbf{e}$
- Can we derive a "similar result" for *2FT*s?
- The answer is yes by incorporating the notion of 'patterns' of computations!

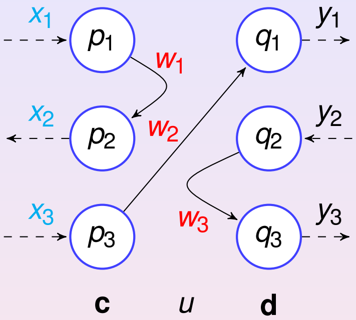
Patterns



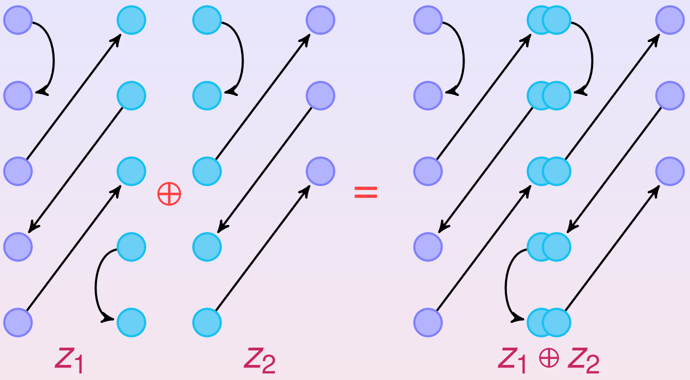
- $p_1 \to p_2$: a right U-turn.
- $p_3 \to q_1$: a right traversal.
- $q_2 \to q_3$: a left U-turn.
- z: (right U-turn, right traversal, left U-turn).

Different Patterns between two Crossing Sequences

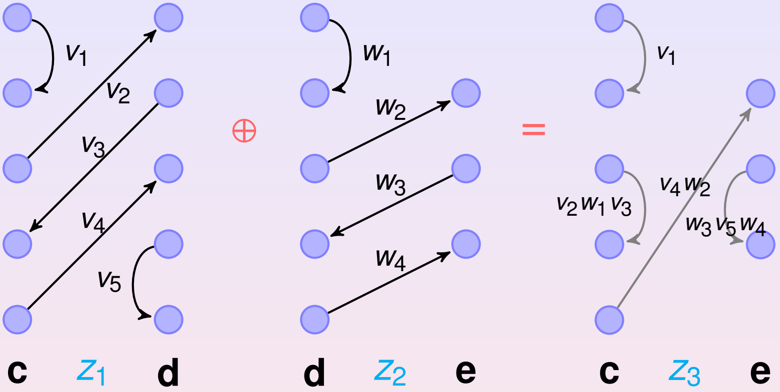




Output: $x_1 v_1 y_1 y_2 v_2 x_2 x_3 v_3 y_3$ Output: $x_1 w_1 x_2 x_3 w_2 y_1 y_2 x_3 y_3$



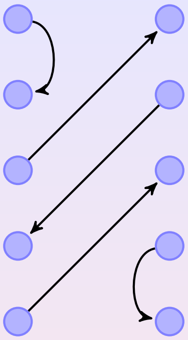
- z_1 : (R-U, R-T, L-T, R-T, L-U)
- z_2 : (R-U, R-T, L-T)
- z_3 : (R-U, R-U, R-T, L-U)
- $\oplus$: an operator (partial mapping) on patterns; e.g. $z_1 \oplus z_2 = z_3$



Lemma

$$\mathbf{c} \xrightarrow[z_1]{u_1|(v_1,\ldots,v_l)_{z_1}} \mathbf{d},\ \mathbf{d} \xrightarrow[z_2]{u_2|(w_1,\ldots,w_m)_{z_2}} \mathbf{e} \Rightarrow \mathbf{c} \xrightarrow[z_1 \oplus z_2]{u_1 u_2|(v_1,\ldots,v_l)_{z_1} \oplus (w_1,\ldots,w_m)_{z_2}} \mathbf{e}.$$

Idempotents



Definition

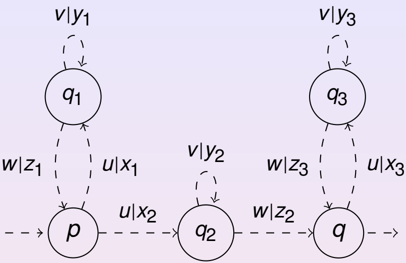
A pattern z is called an idempotent if $z \oplus z = z$.

Example

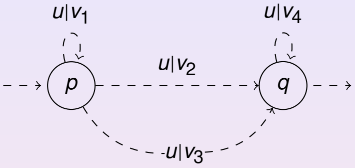
- $z_1 \oplus z_1 \neq z_1$.
- $(z_1 \oplus z_1) \oplus (z_1 \oplus z_1) = (z_1 \oplus z_1)$.

Lemma

If $z \oplus z$ exists, then $z' = \underbrace{z \oplus \cdots \oplus z}_{p}$ for some p and z' is an idempotent.



IV1: $|y_1| \neq |y_2|$



IV2: $v_1 \neq \epsilon$ and $\exists i, v_2(i) \neq v_3(i)$

- Can we derive a "crossing sequence version" of criteria for infinite-valued *2FTs*?
- Perhaps by replacing states and strings with crossing sequences and vectors of strings, respectively?

Length-conflicts and Position-conflicts

Given $v_1, v_2 \in \Gamma^*$ with $v_1 \neq v_2$, one of the following statements is true:

1. $|v_1| \neq |v_2|$. (Length-conflict)
2. $v_1(i) \neq v_2(i)$ for some i. (Position-conflict)

Length-conflicts and Position-conflicts

Given $v_1, v_2 \in \Gamma^*$ with $v_1 \neq v_2$, one of the following statements is true:

1. $|v_1| \neq |v_2|$. (Length-conflict)
2. $v_1(i) \neq v_2(i)$ for some i. (Position-conflict)

Given a 2FT T:

Length-conflict:

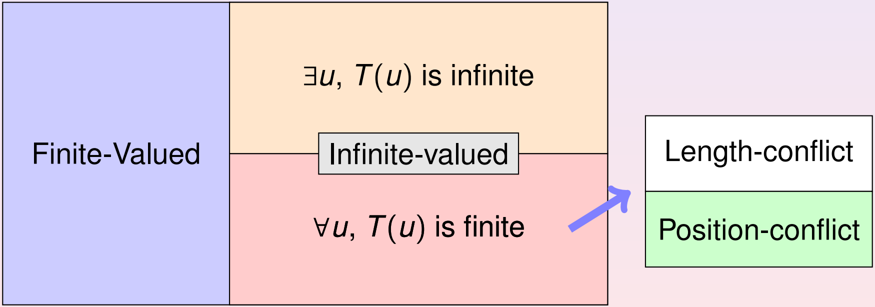
$\forall n$, there are $u, v_1, \ldots, v_n$ such that $\forall i, (u, v_i) \in R(T)$ and $\forall i \neq j, v_i$ and v_j have a length-conflict.

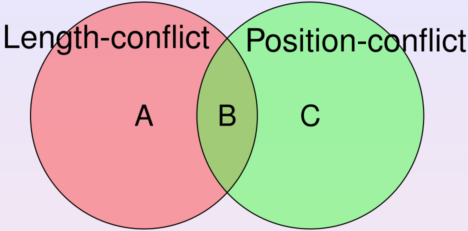
Position-conflict:

$\forall n$, there are $u, v_1, \ldots, v_n$ such that $\forall i, (u, v_i) \in R(T)$ and $\forall i \neq j, v_i$ and v_j have a position-conflict.

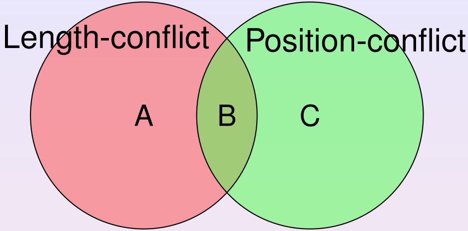
Proposition

A 2FT T is infinite-valued iff T has length-conflicts or position-conflicts.

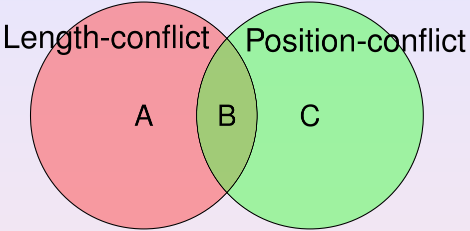




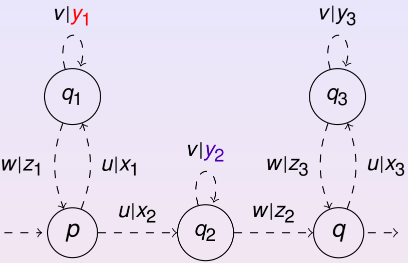
- For *1FTs*,
  $A \cup B \Rightarrow$ (IV1) $\Leftrightarrow$ Length-conflict.
  $C \Rightarrow$ (IV2) $\Leftrightarrow$ Position-conflict.



- For *1FTs*,
  $A \cup B \Rightarrow$ (IV1) $\Leftrightarrow$ Length-conflict.
  $C \Rightarrow$ (IV2) $\Leftrightarrow$ Position-conflict.
- Can we derive similar results for *2FTs*?



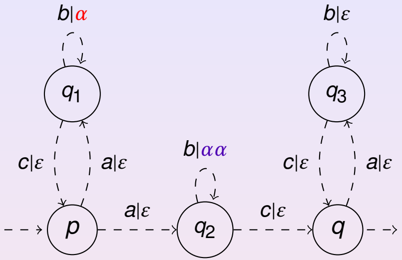
- For *1FTs*,
  $A \cup B \Rightarrow$ (IV1) $\Leftrightarrow$ Length-conflict.
  $C \Rightarrow$ (IV2) $\Leftrightarrow$ Position-conflict.
- Can we derive similar results for *2FTs*?
- The answer is yes. But, while considering the 'position-conflict' case, a more subtle argument is needed!



IV1: $|y_1| \neq |y_2|$

Theorem (Weber, 90)

A *1FT T* has length-conflicts iff it satisfies criterion IV1.

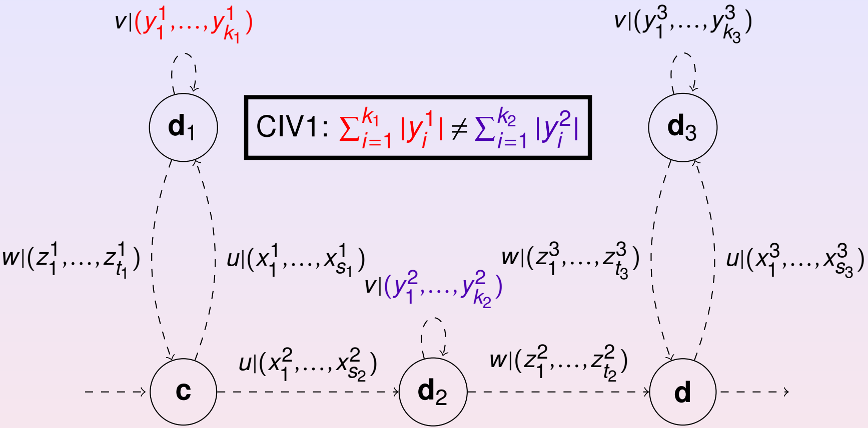


Theorem (Weber, 90)

A *1FT T* has length-conflicts iff it satisfies criterion IV1.

Example

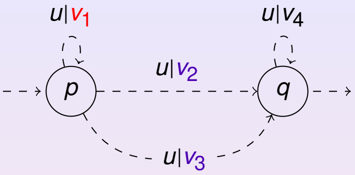
- $|\alpha| \neq |\alpha\alpha|$.
- On $(abc)(ab^2c)\ldots(ab^nc)$, there are n different outputs: $\alpha^2, \alpha^5, \ldots, \alpha^{i+2(i+1)}, \ldots,$ and α^{3n-1}.



Theorem

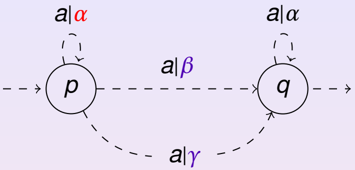
A 2FT T has length-conflicts iff it satisfies criterion CIV1.

- (IV2) is a sufficient condition but not a necessary condition for position-conflicts.
- For *1FTs* without length-conflicts, (IV2) is a necessary and sufficient condition for position-conflicts.



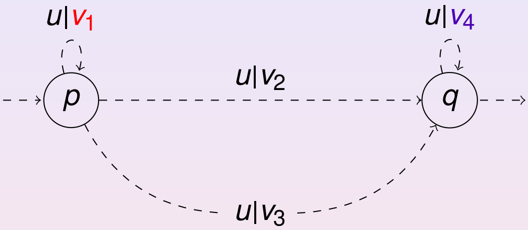
IV2: $v_1 \neq \epsilon$ and $\exists i, v_2(i) \neq v_3(i)$

- (IV2) is a sufficient condition but not a necessary condition for position-conflicts.
- For *1FT*s without length-conflicts, (IV2) is a necessary and sufficient condition for position-conflicts.



Example

- $|\alpha| \neq 0$.
- β and γ have a position-conflict.
- On a^n, there are $2n$ different outputs, $\beta\alpha^{n-1}, \gamma\alpha^{n-1}, \alpha\beta\alpha^{n-2}, \alpha\gamma\alpha^{n-2}, \ldots, \alpha^{n-1}\beta$, and $\alpha^{n-1}\gamma$.

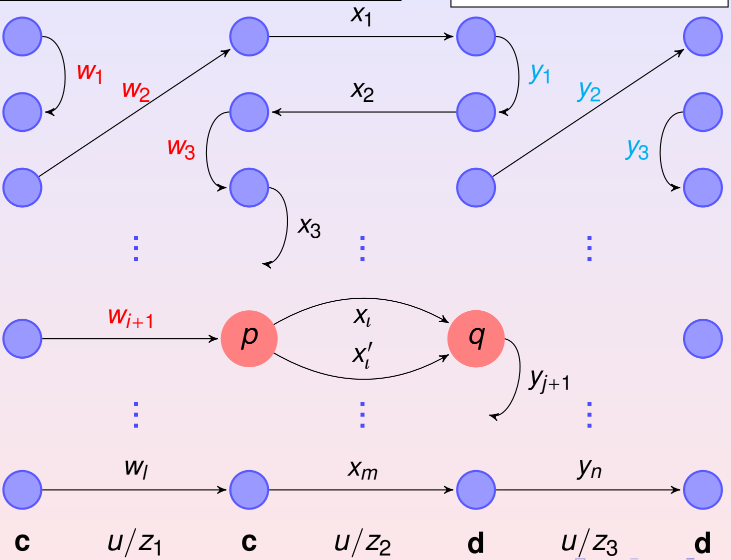


IV2: $v_1 \neq \epsilon$ and $\exists i, v_2(i) \neq v_3(i)$

CIV2: Non-empty String Modification

- x_ι and x'_ι have a position-conflict

- $\sum_{t=1}^{i+1} |w_t| \neq \sum_{t=1}^{j} |y_t|$



There exists an accepting computation π along which there are crossing sequences $\mathbf{c}, \mathbf{d}$, patterns z_1, z_2, z_3, strings $u \in \Sigma^+$, $w_1, \ldots, w_l$, $x_1, \ldots, x_m$, $x'_1, \ldots, x'_m$, $y_1, \ldots, y_n \in \Gamma^*$,

1. (The Original)
   - $\mathbf{c} \xrightarrow[z_1]{u|(w_1,\ldots,w_l)_{z_1}} \mathbf{c}$, • $\mathbf{c} \xrightarrow[z_2]{u|(x_1,\ldots,x_m)_{z_2}} \mathbf{d}$,
   - $\mathbf{c} \xrightarrow[z_2]{u|(x'_1,\ldots,x'_m)_{z_2}} \mathbf{d}$, • $\mathbf{d} \xrightarrow[z_3]{u|(y_1,\ldots,y_n)_{z_3}} \mathbf{d}$,

   and for some index $\iota \in [m]$, there is a position-conflict between x_ι and x'_ι
2. (The Modified)
   x_ι (x'_ι) corresponds to either a left or a right traversal between some state p in $\mathbf{c}$ and some state q in $\mathbf{d}$, and
   - (right traversal): if $\pi : \cdots\ p_1 \xrightarrow{u|w_j} p \xrightarrow{u|x_\iota} q \xrightarrow{u|y_k} q_1\ \cdots$, then $\sum_{t=1}^{j} |w_t| \neq \sum_{t=1}^{k-1} |y_t|$
   - (left traversal): if $\pi : \cdots\ q_1 \xrightarrow{u|y_k} q \xrightarrow{u|x_\iota} p \xrightarrow{u|w_j} p_1\ \cdots$, then $\sum_{t=1}^{j-1} |w_t| \neq \sum_{t=1}^{k} |y_t|$

Theorem

Let T be a bounded-crossing 2FT having no length-conflicts. T is infinite-valued iff it satisfies CIV2.

Proof.

($\Leftarrow$) The 'only if' direction can be derived easily.

($\Rightarrow$) The 'if' direction is divided into the following steps:

1. Lemma A
2. Two base cases
3. General cases

$\square$

Lemma A

Given an infinite-valued *2FT* *T* without length-conflicts, then

1. $\mathbf{c} \xrightarrow[z_1]{u|(w_1,\ldots,w_l)_{z_1}} \mathbf{c}$,
2. $\mathbf{c} \xrightarrow[z_2]{u|(x_1,\ldots,x_m)_{z_2}} \mathbf{d}$,
3. $\mathbf{c} \xrightarrow[z_2]{u|(x'_1,\ldots,x'_m)_{z_2}} \mathbf{d}$,
4. $\mathbf{d} \xrightarrow[z_3]{u|(y_1,\ldots,y_n)_{z_3}} \mathbf{d}$,
5. $|x_i| = |x'_i|, 1 \leq i \leq m$, x_ι and x'_ι have a position conflict, for some ι,
6. $|x_\iota| = |x'_\iota| > \psi$, and
7. z_1 and z_3 are idempotents.

for some $u \in \Sigma^+$, ect.

Two Base Cases

1. All properties in Lemma A are satisfied.
2. $p_2 = p_1 = p$.
3. $q_2 = q_1 = q$.
4. Computations from p' to p_1 and from q_1 to q' are sequences of U-turns.

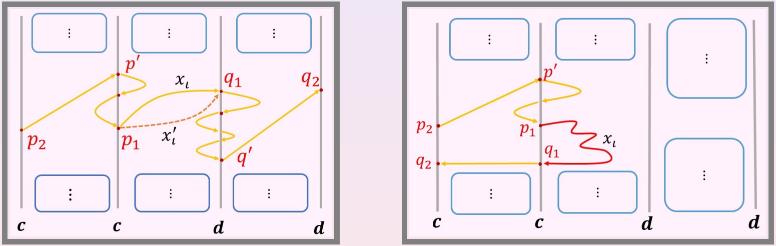
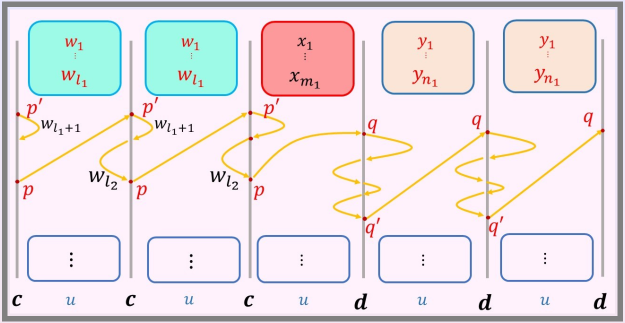


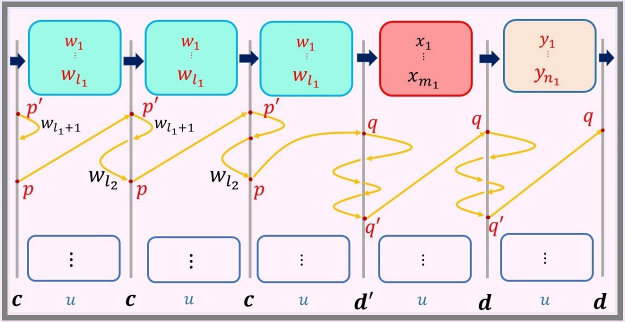
Figure: Base case: traversal (left); U-turn (right).

Base Case: Traversal



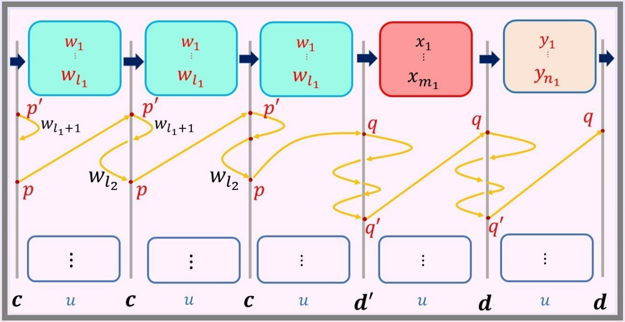
- Concatenate computation $\mathbf{c} \xrightarrow[z_1]{u|\vec{w}} \mathbf{c}$ and $\mathbf{d} \xrightarrow[z_3]{u|\vec{y}} \mathbf{d}$ to the left and to the right of the original computation. (Since $\mathbf{c} = \mathbf{c}$ and $\mathbf{d} = \mathbf{d}$.)

Base Case: Traversal



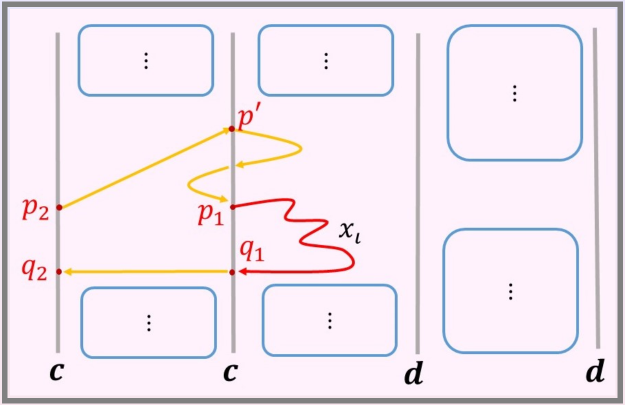
- Concatenate computation $\mathbf{c} \xrightarrow[z_1]{u|\vec{w}} \mathbf{c}$ and $\mathbf{d} \xrightarrow[z_3]{u|\vec{y}} \mathbf{d}$ to the left and to the right of the original computation. (Since $\mathbf{c} = \mathbf{c}$ and $\mathbf{d} = \mathbf{d}$.)
- Shift.

Base Case: Traversal



- Since T has no length-conflicts,
  - $\sum_{i=1}^{l_1} |w_i| = \sum_{i=1}^{n_1} |y_i|$ and
  - $\sum_{i=l_1+1}^{l_2} |w_i| \neq 0$.
- Therefore, $\sum_{i=1}^{l_2} |w_i| > \sum_{i=1}^{n_1} |y_i|$. ($\Rightarrow$ CIV2 - The Modified)

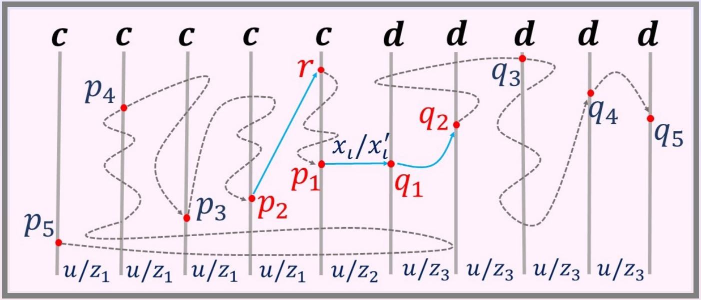
Base Case: U-turn



- Does not exist!

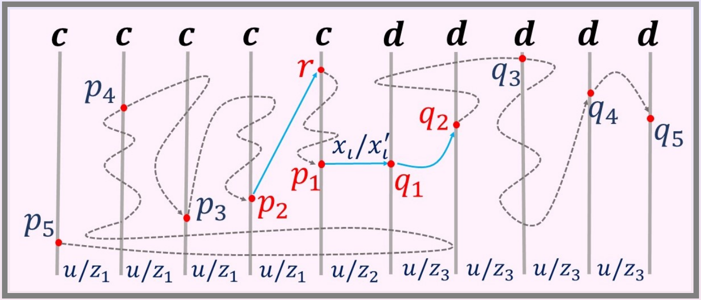
General Case

- $p_2 \neq p_1$



General Case

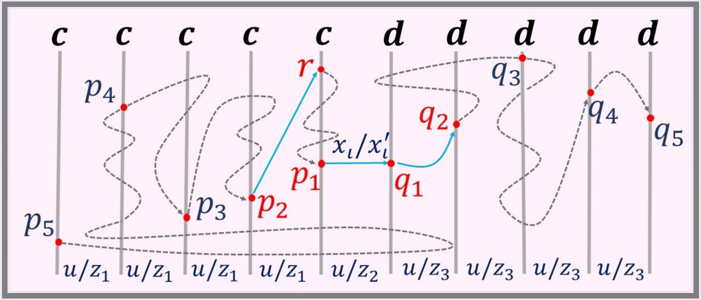
- $p_2 \neq p_1$



1. Keep concatenating computation $\mathbf{c} \xrightarrow[z_1]{u|\vec{w}} \mathbf{c}$ and $\mathbf{d} \xrightarrow[z_3]{u|\vec{y}} \mathbf{d}$ to the left and to the right of the original computation.
2. Then, $p_i = p_2$ and $q_i = q_2$, for each $i > 2 \Rightarrow$ Base Cases.

General Case

- $p_2 \neq p_1$



Lemma

If z_1 and z_3 are idempotents and $z_1 \oplus z_2 \oplus z_3$ exists, then
$$\underbrace{z_1 \oplus \cdots \oplus z_1}_{n} \oplus z_2 \oplus \underbrace{z_3 \oplus \cdots \oplus z_3}_{n} = z_1 \oplus z_2 \oplus z_3,$$ *for all n.*

Future Work

1. Decidability of finite-valuedness:
   We surmise that the problem is decidable, perhaps through a detailed analysis of our techniques.
2. Decomposition of finite-valued *2FT*s.
   In view of the decomposability result of finite-valued 1FTs, can similar results be obtained?
3. One-way definability of finite-valued *2FT*s.
4. Finite-valuedness of streaming string transducers.
   ($D2FT = DSST = MSO$)

Thank You for Your Time!