Regular Grammars Generate Regular Languages

Theorem

Regular grammars generate exactly the class of regular languages:

If G is a regular grammar then $L(G)$ is a regular language

If L is a regular language then there is a regular grammar G with $L(G) = L$

Proof

First we prove:

If G is a regular grammar then $L(G)$ is a regular language

G can be:

Right-linear grammar

or

Left-linear grammar

The case of Right-Linear Grammars

Let G be a right-linear grammar

We will show: $L(G)$ is regular

The proof:

We will construct an NFA M

with $L(M) = L(G)$

Grammar G is right-linear

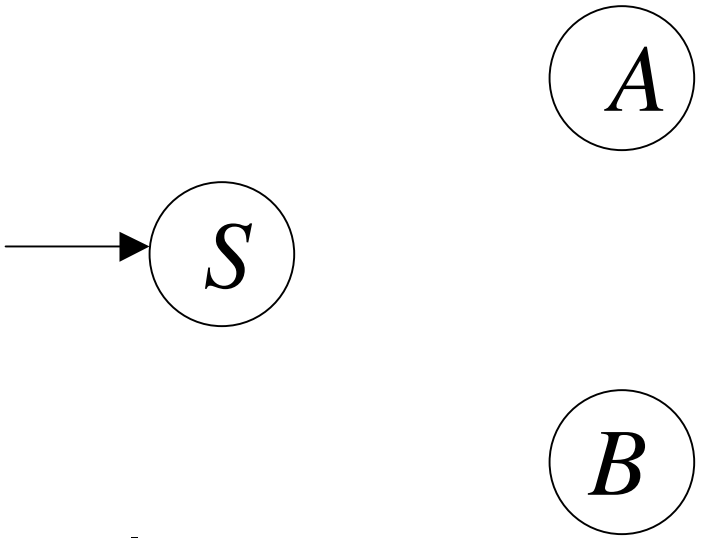
Example:

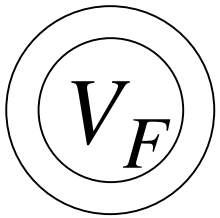
$$S \to aA \mid B$$

$$A \to aa\ B$$

$$B \to b\ B \mid a$$

Construct NFA M such that every state is a variable:





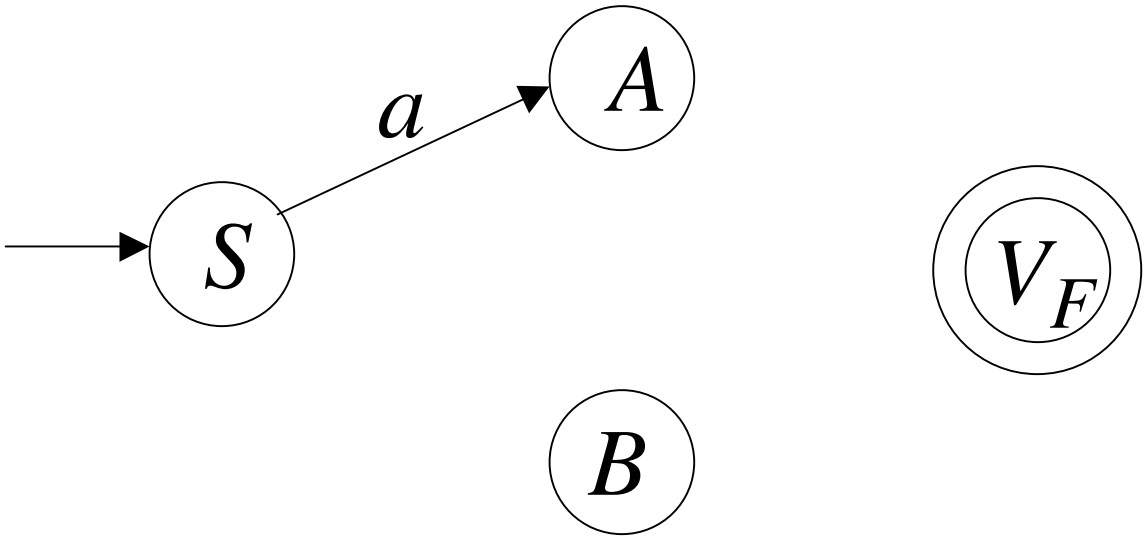
special final state

$$S \to aA \,|\, B$$

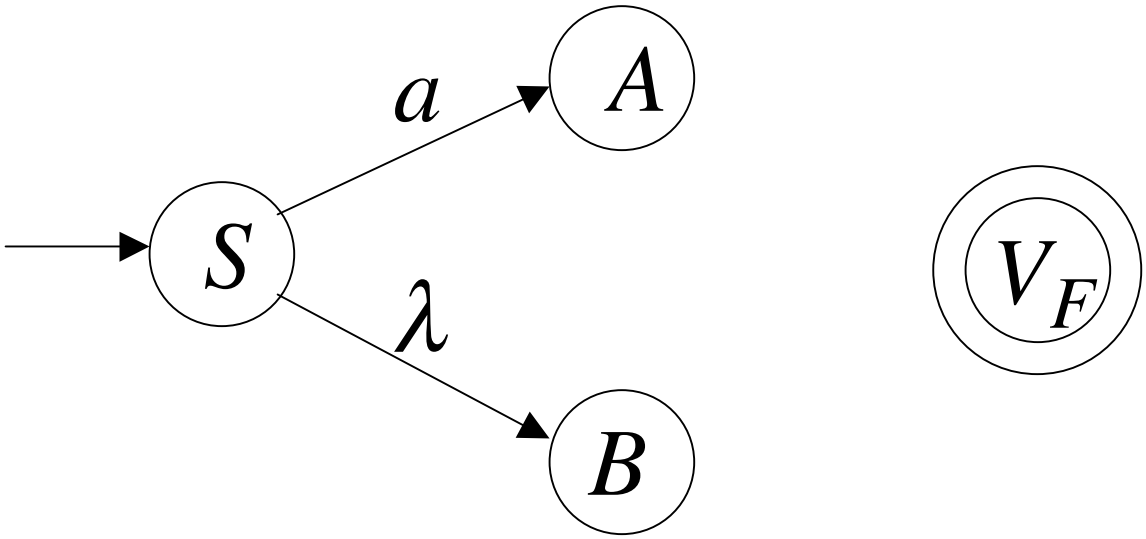
$$A \to aa\ B$$

$$B \to b\ B \,|\, a$$

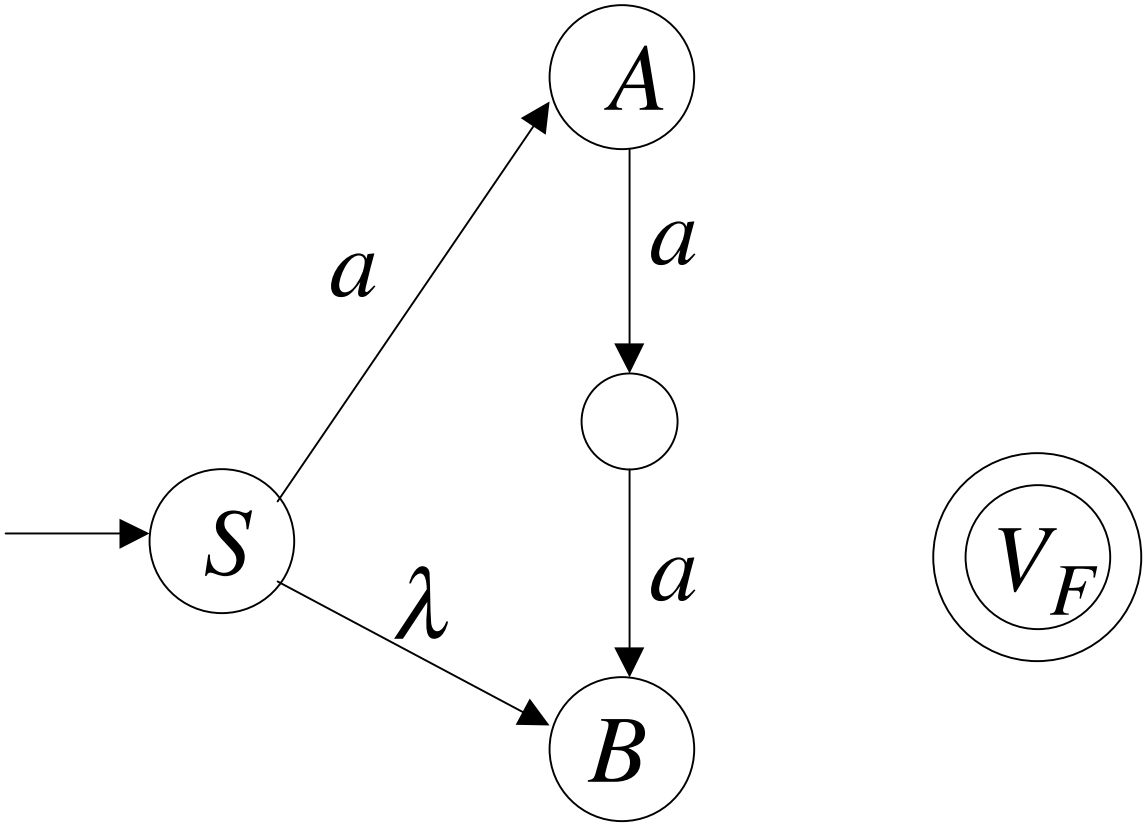
Add edges for each production:



$$S \rightarrow aA$$

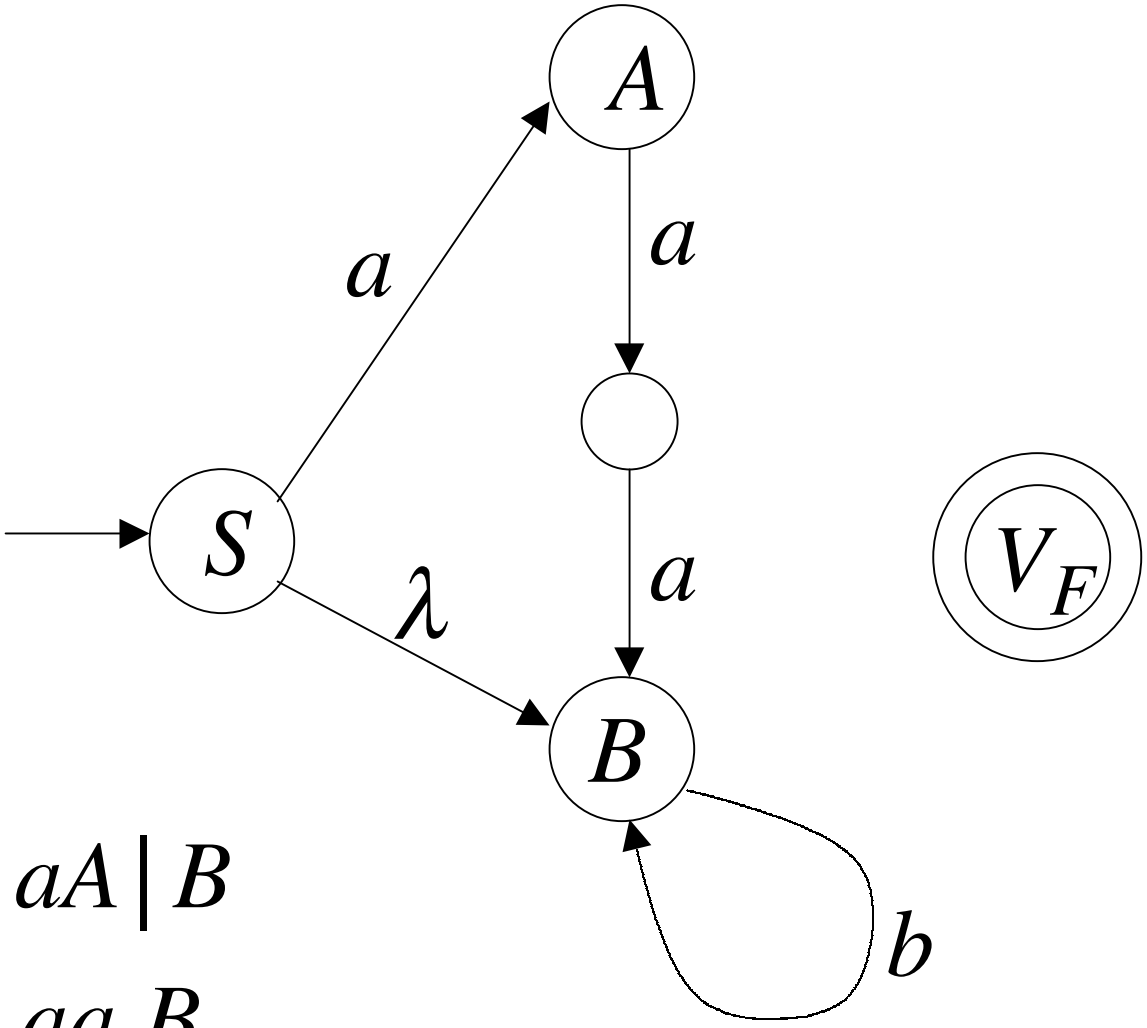


$$S \rightarrow aA \mid B$$



$$S \rightarrow aA \mid B$$

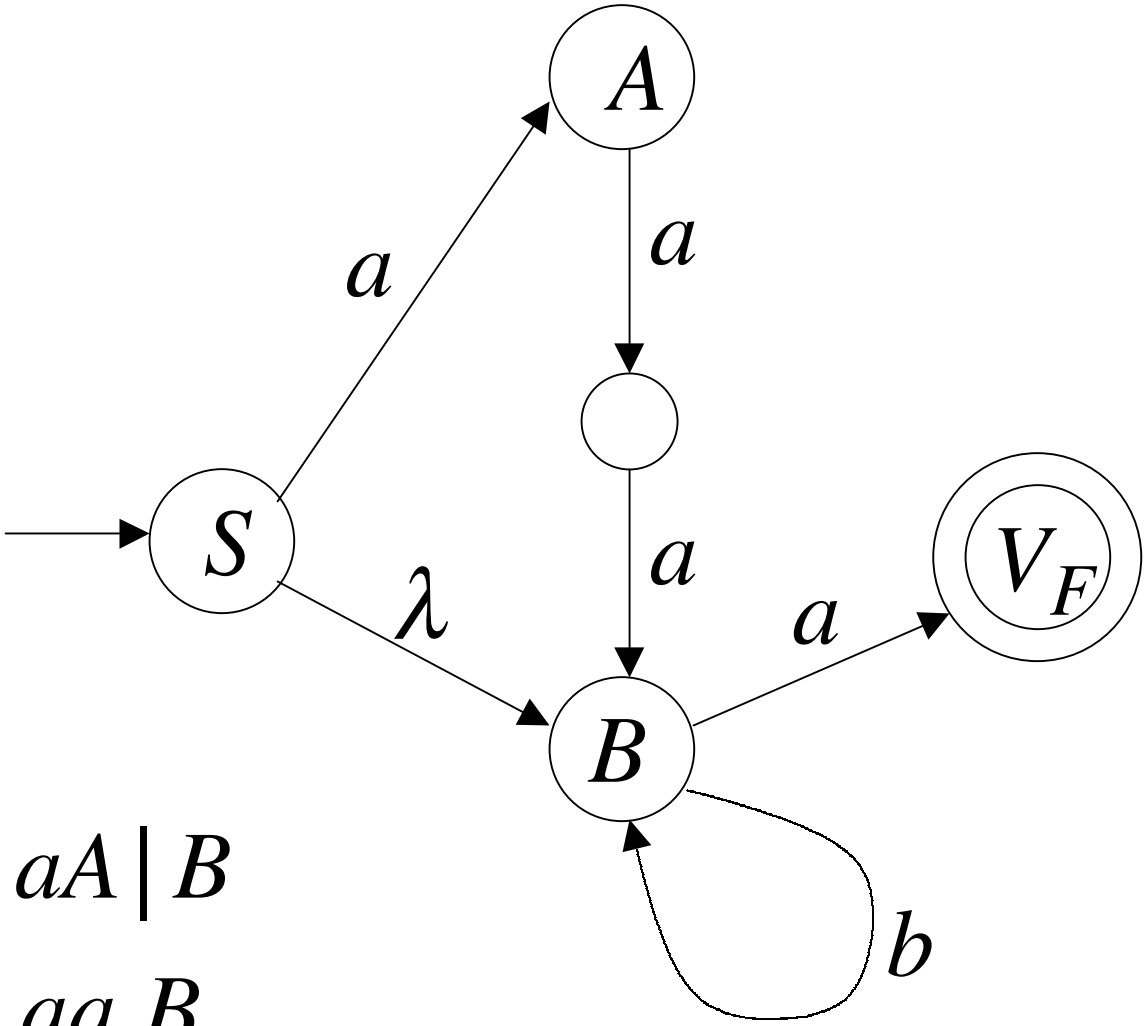
$$A \rightarrow aa\ B$$



$$S \rightarrow aA \mid B$$

$$A \rightarrow aa\ B$$

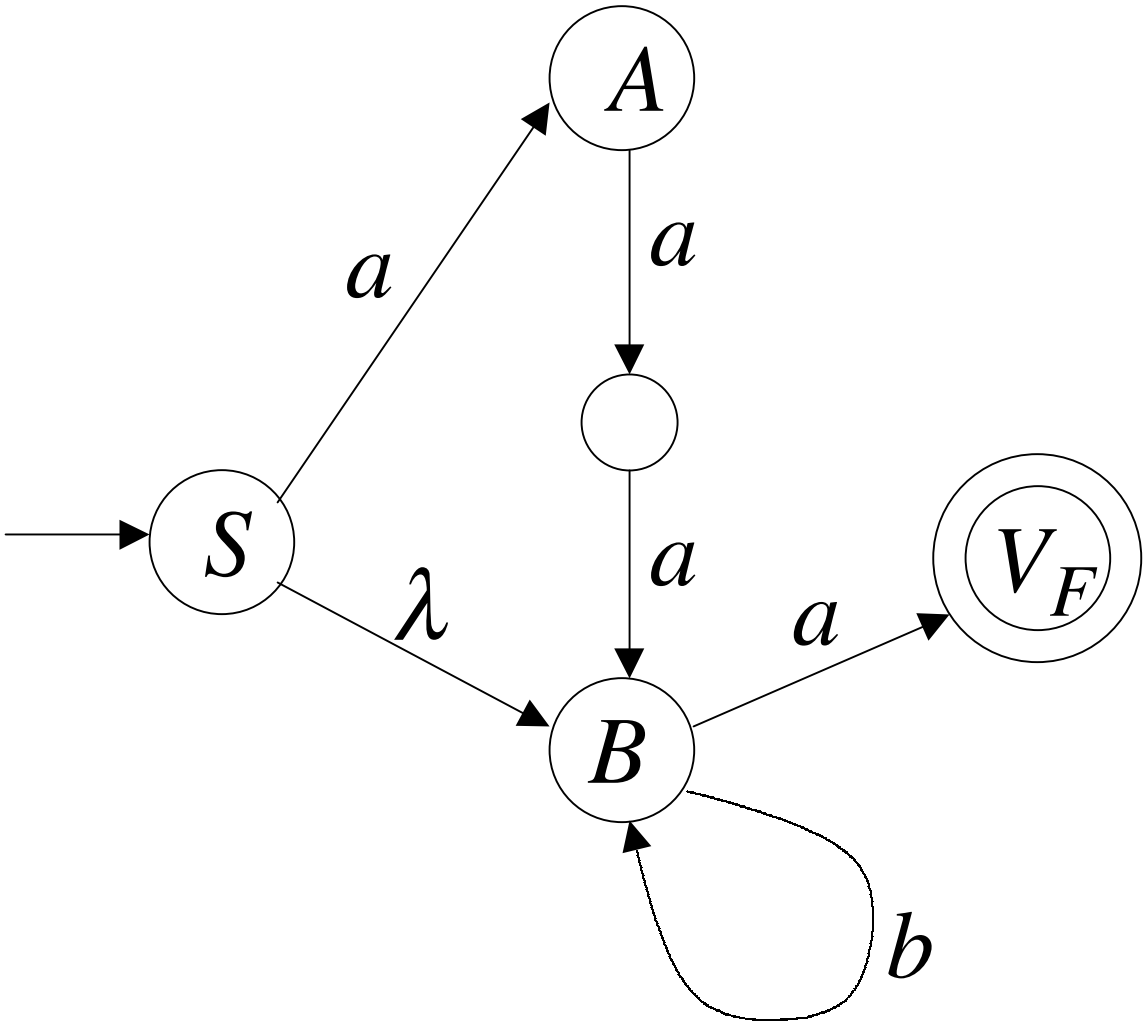
$$B \rightarrow bB$$



$$S \rightarrow aA \mid B$$

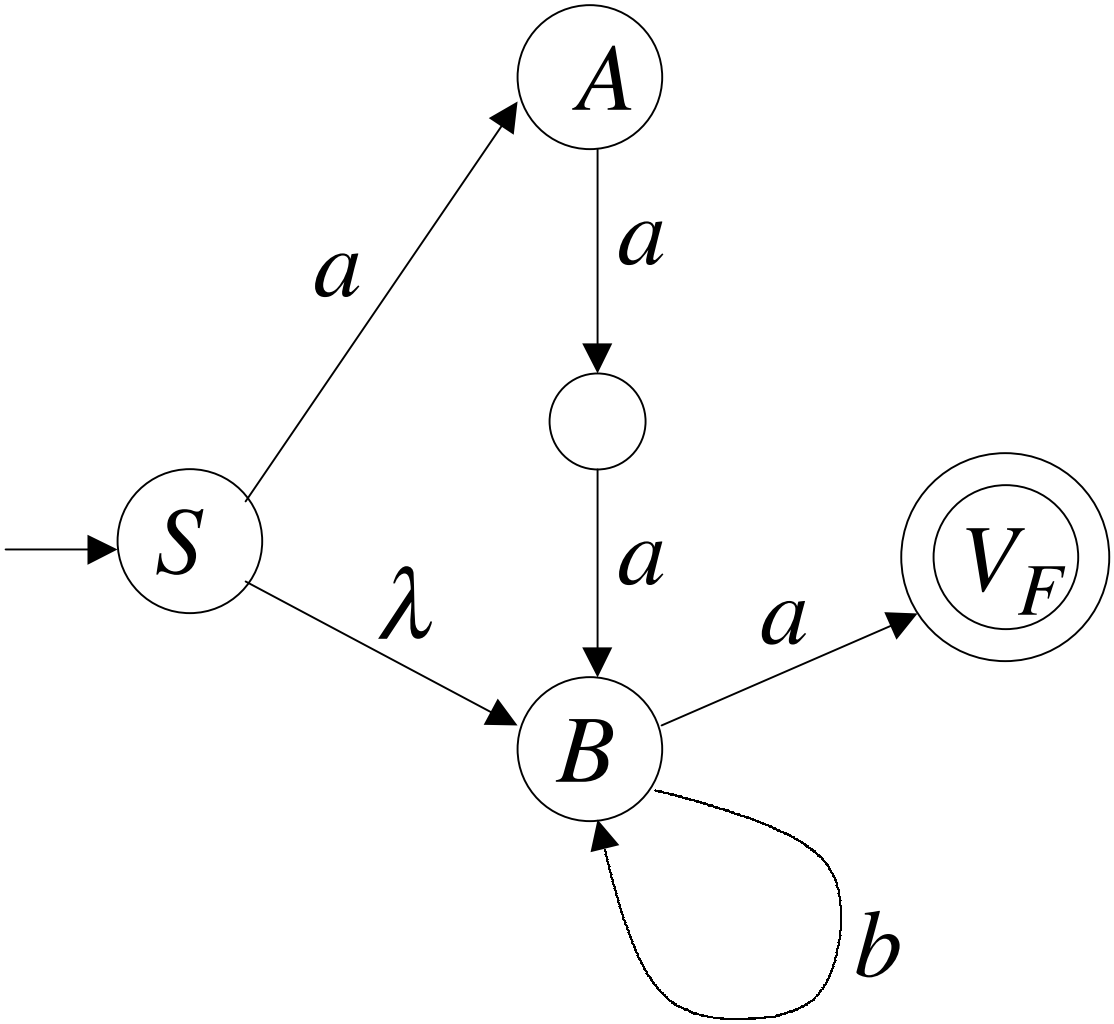
$$A \rightarrow aa\ B$$

$$B \rightarrow bB \mid a$$



$$S \Rightarrow aA \Rightarrow aaaB \Rightarrow aaabB \Rightarrow aaaba$$

NFA M



Grammar

G

$S \to aA \mid B$

$A \to aa\ B$

$B \to bB \mid a$

$L(M) = L(G) =$

$aaab^* a + b^* a$

In General

A right-linear grammar G

has variables: $V_0, V_1, V_2, \ldots$

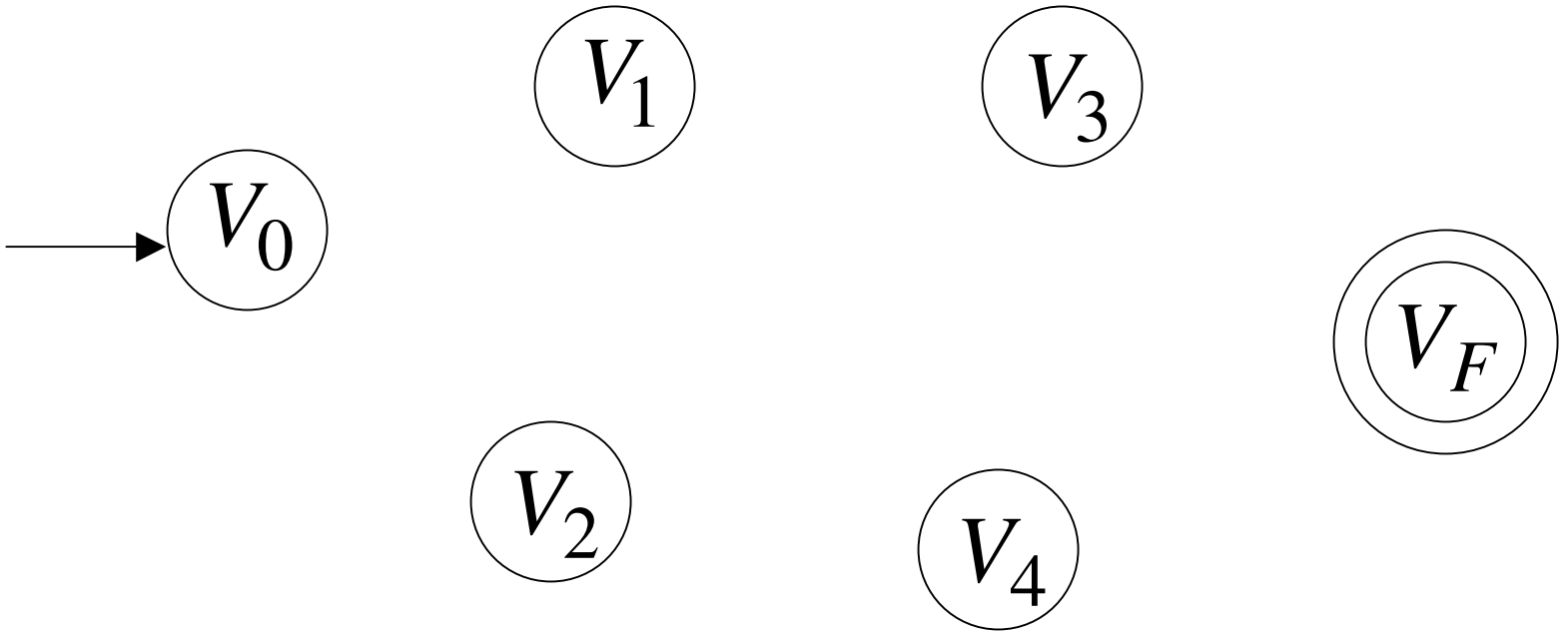
and productions: $V_i \to a_1 a_2 \cdots a_m V_j$

or

$V_i \to a_1 a_2 \cdots a_m$

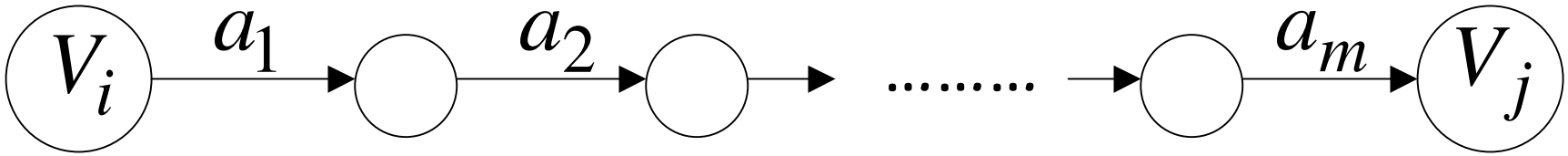
We construct the NFA M such that:

each variable V_i corresponds to a node:



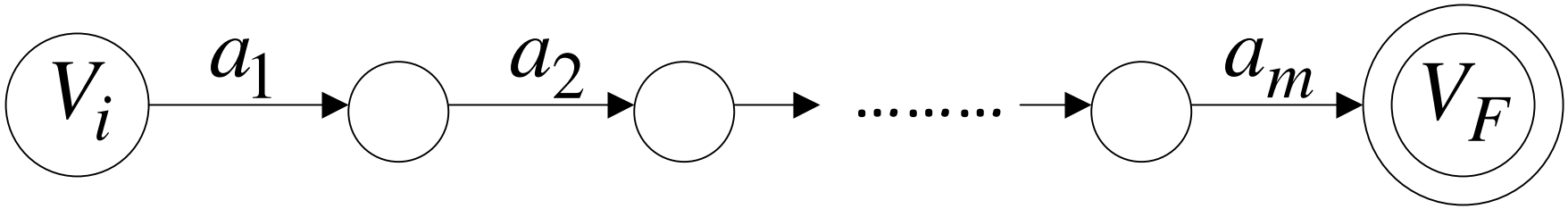
For each production: $V_i \to a_1 a_2 \cdots a_m V_j$

we add transitions and intermediate nodes

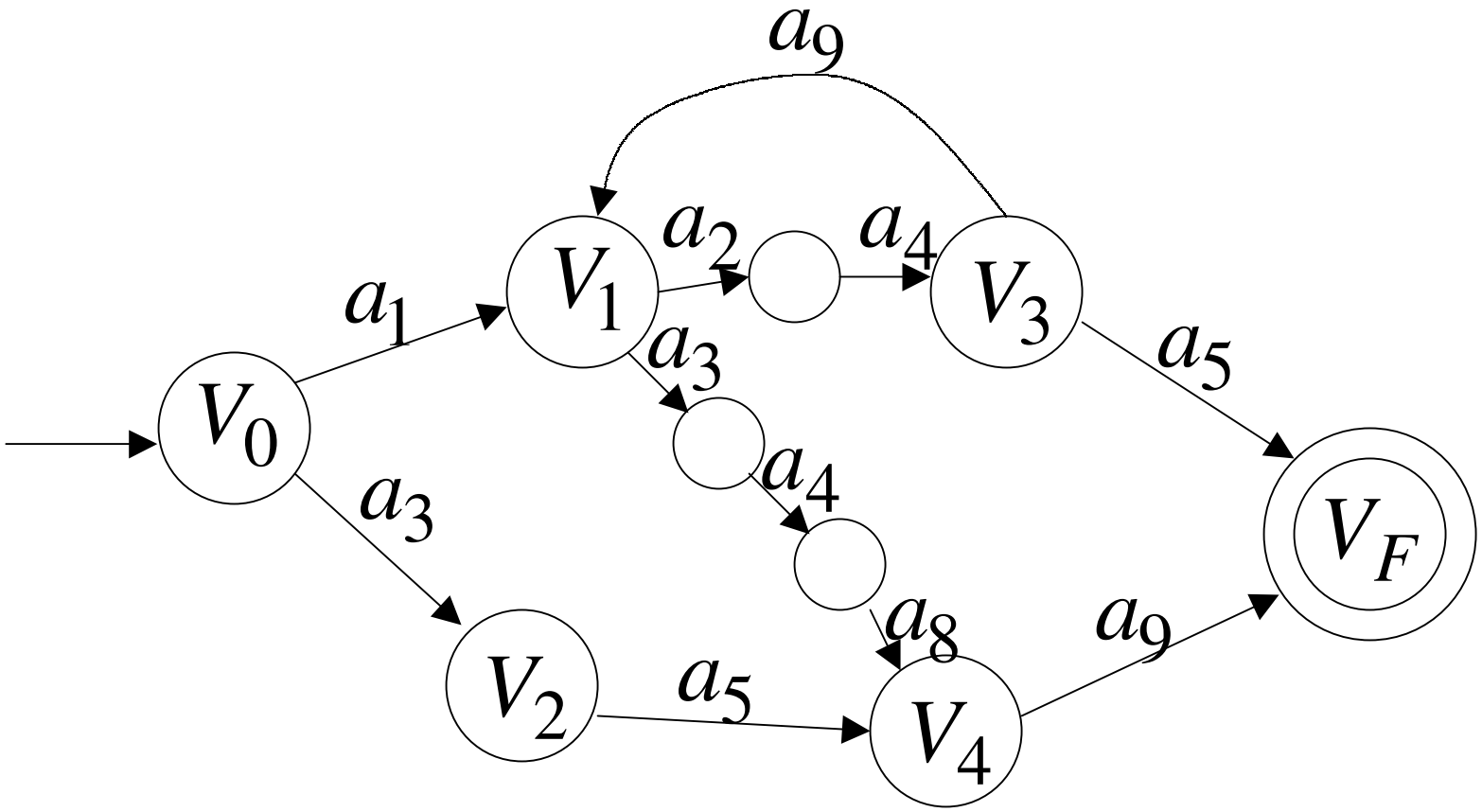


For each production: $V_i \rightarrow a_1 a_2 \cdots a_m$

we add transitions and intermediate nodes



Resulting NFA M looks like this:



Now, we need to show:

$$L(G) = L(M)$$

$$L(G) \subseteq L(M)$$

$$w \in L(G) \Rightarrow w \in L(M)$$

$$L(M) \subseteq L(G)$$

$$w \in L(M) \Rightarrow w \in L(G)$$

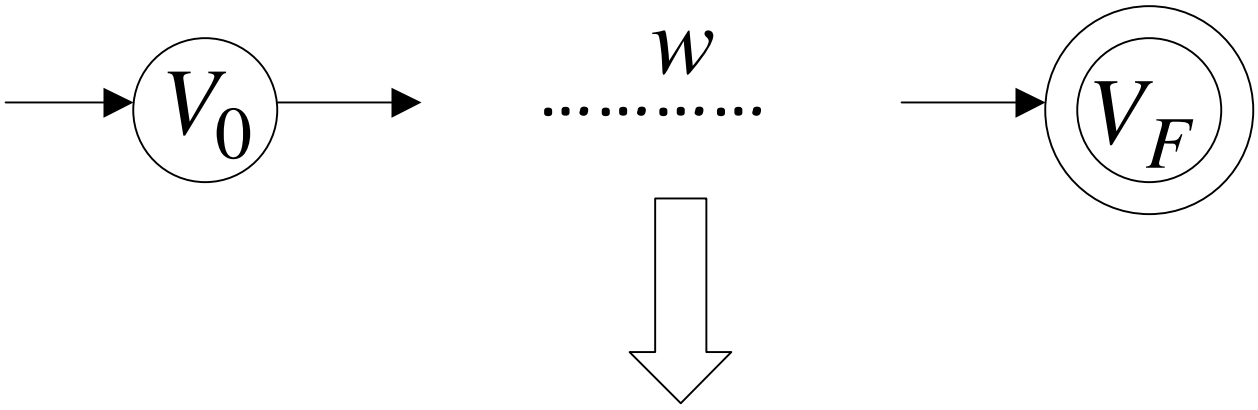
The Case: $L(G) \subseteq L(M)$

Take $w \in L(G)$

We will show:

there is a path with label w in M

from state V_0 to state V_f



$w \in L(M)$

Grammar G looks like:

$$V_0 \to xV_1$$
$$V_1 \to yV_2$$
$$\ldots\ldots$$
$$V_n \to z$$

strings

$$x, y, \ldots, z$$

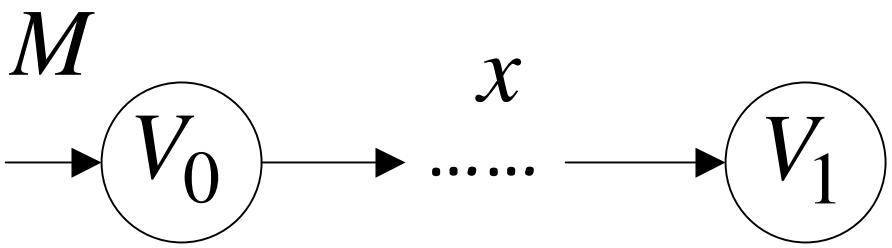
$$w = xy\cdots z$$

$w \in L(G)$:

$$V_0 \Rightarrow xV_1 \Rightarrow xyV_2 \Rightarrow \quad \ldots \quad \Rightarrow xy\cdots V_n \Rightarrow xy\cdots z$$

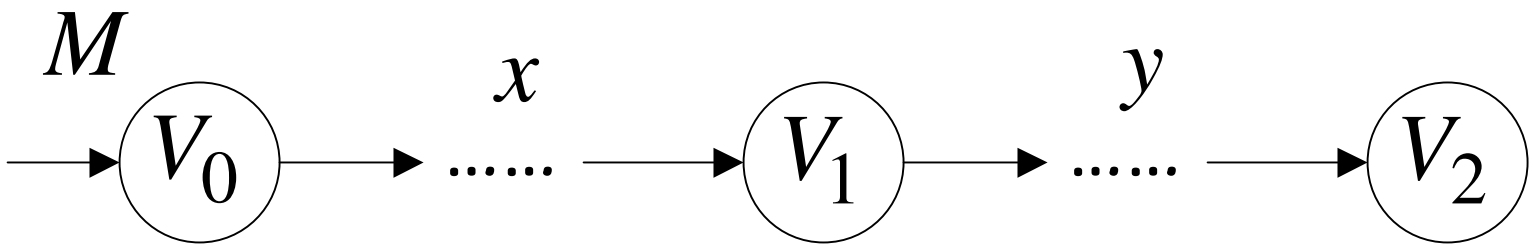
$$V_0 \to xV_1$$
$$V_1 \to yV_2$$
$$\ldots\ldots$$
$$V_n \to z$$

$$V_0 \Rightarrow xV_i$$



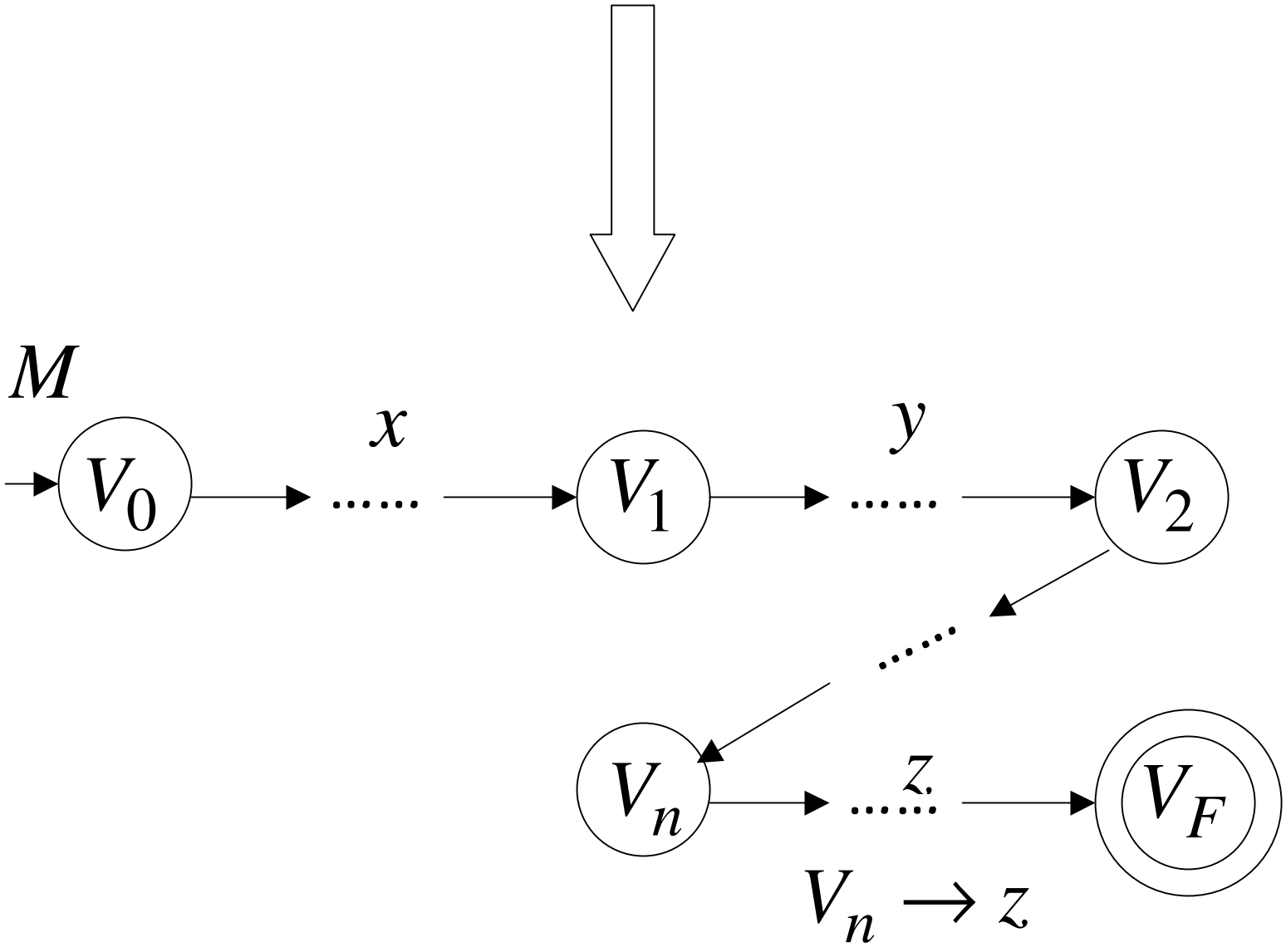
$$V_0 \rightarrow xV_i$$

$$V_0 \Rightarrow xV_i \Rightarrow xyV_z$$

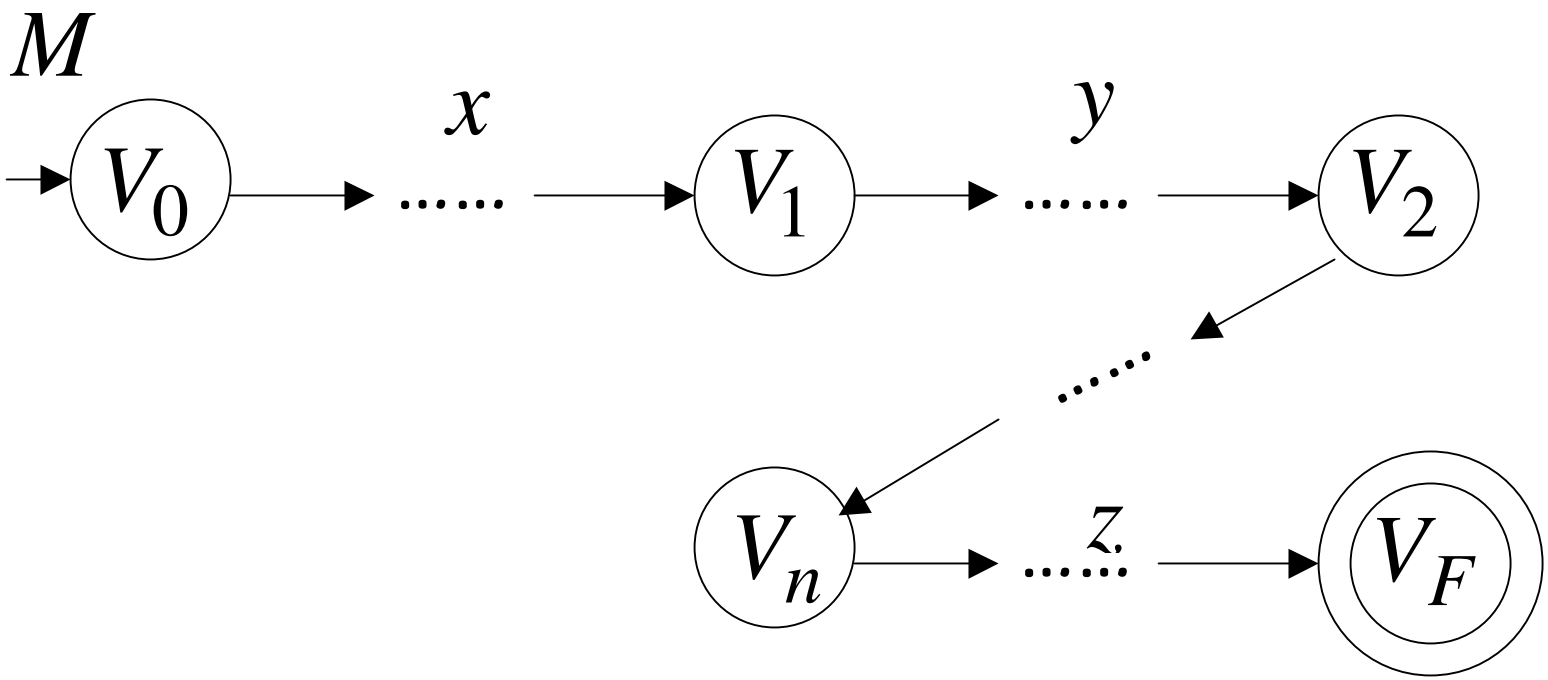


$$V_1 \rightarrow yV_2$$

$$V_0 \Rightarrow xV_1 \Rightarrow xyV_2 \Rightarrow \quad \ldots \quad \Rightarrow xy\cdots V_n \Rightarrow xy\cdots z$$



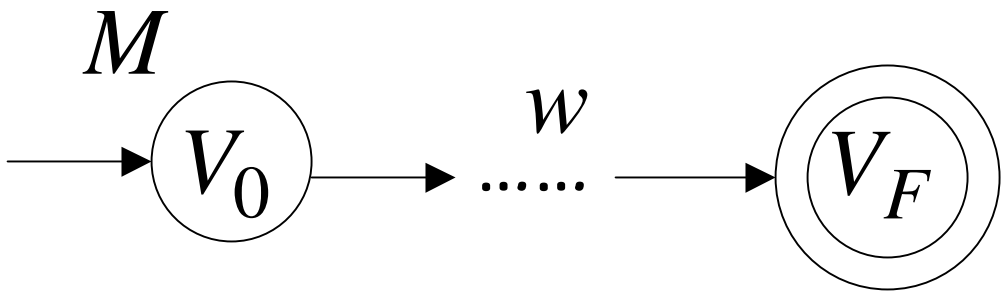
Since:



We have: $xy \cdots z \in L(M)$

Since: $w = xy\cdots z$ and $xy\cdots z \in L(M)$

We have:

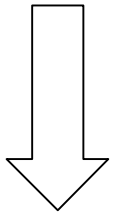


We have: $w \in L(M)$

The Case: $L(M) \subseteq L(G)$

Take $w \in L(M)$

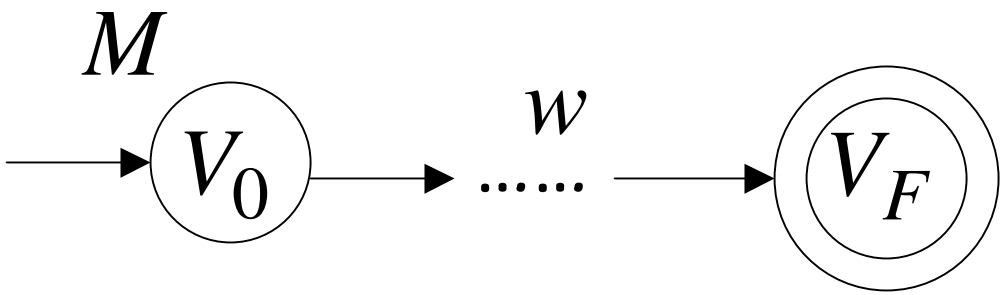
We will show that in G: $V_0 \overset{*}{\Rightarrow} w$



$w \in L(G)$

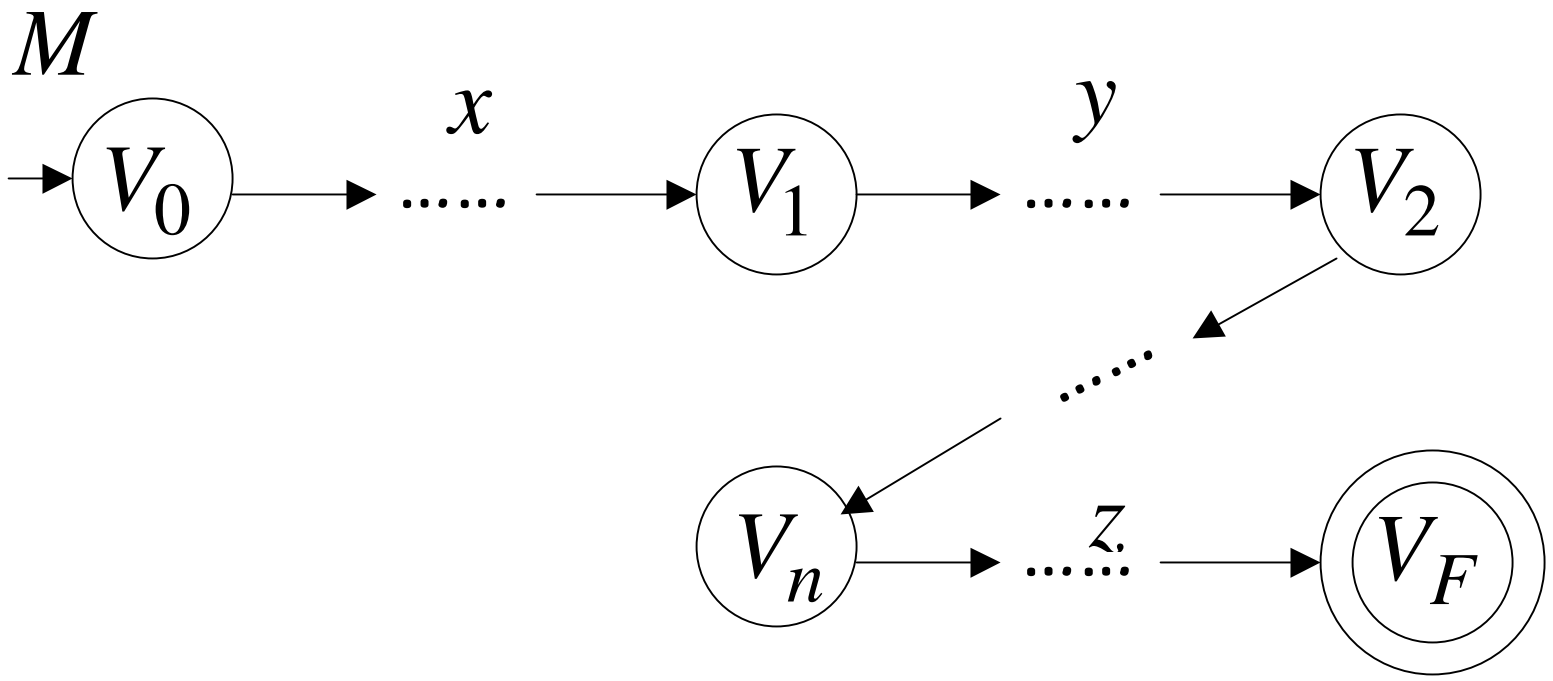
Since $w \in L(M)$

there is a path

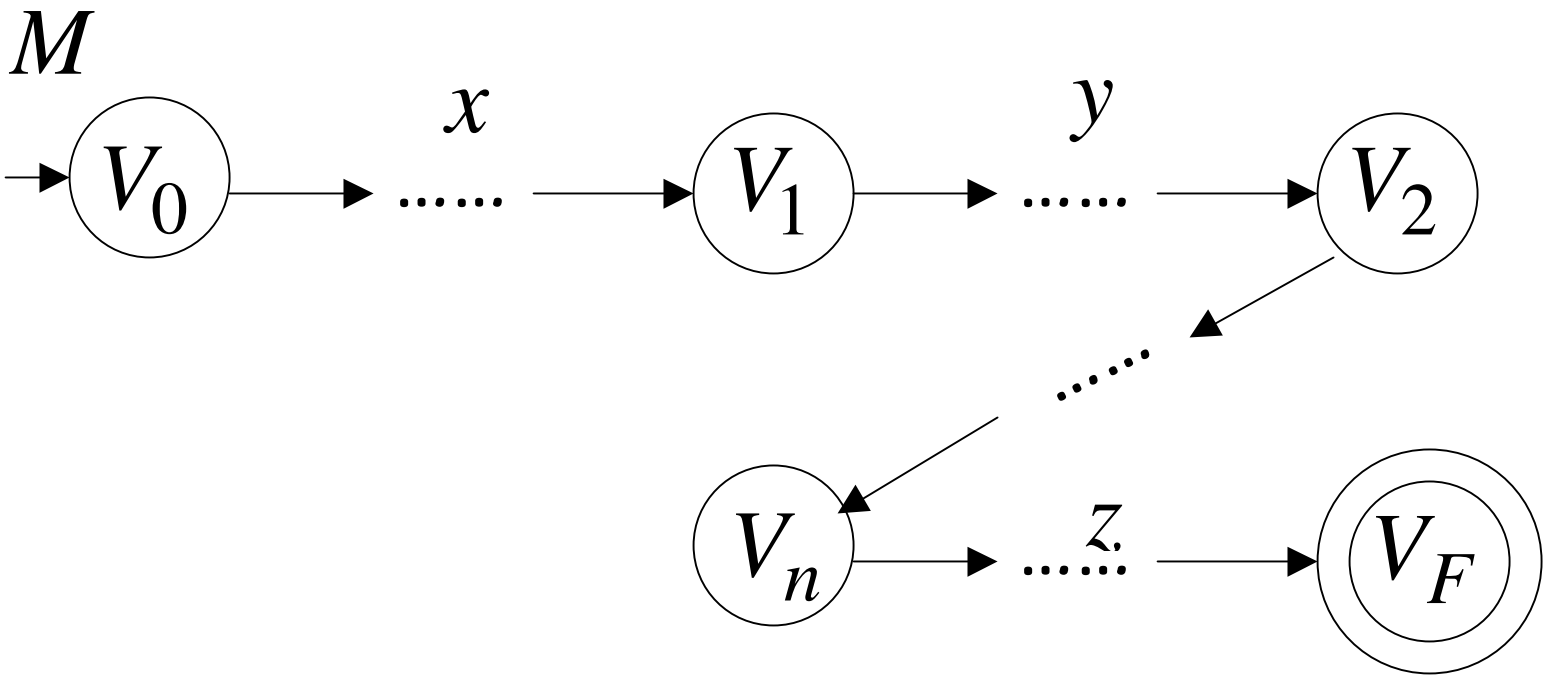


Write: $w = xy \cdots z$

There is a path



Since:



This derivation is possible

$$V_0 \Rightarrow xV_1 \Rightarrow xyV_2 \Rightarrow \quad \ldots \quad \Rightarrow xy\cdots V_n \Rightarrow xy\cdots z$$

Since:

$$V_0 \Rightarrow xV_1 \Rightarrow xyV_2 \Rightarrow \quad \ldots \quad \Rightarrow xy\cdots V_n \Rightarrow xy\cdots z$$

$$w = xy\cdots z$$

We have: $V_0 \overset{*}{\Rightarrow} w$

We have: $w \in L(G)$

The Case of Left-Linear Grammars

Let G be a left-linear grammar

We will show: $L(G)$ is regular

The proof:

We will construct a right-linear grammar G' with $L(G) = (L(G'))^R$

Since G is left-linear grammar the productions look like:

$$A \to a_1 a_2 \cdots a_k B$$

$$A \to a_1 a_2 \cdots a_k$$

Construct right-linear grammar G'

In G :

$$A \to a_1 a_2 \cdots a_k B$$

$$A \to vB$$

In G' :

$$A \to a_k \cdots a_2 a_1 B$$

$$A \to v^R B$$

In G :

$$A \to a_1 a_2 \cdots a_k$$

$$A \to v$$

In G' :

$$A \to a_k \cdots a_2 a_1$$

$$A \to v^R$$

It is easy to see that: $L(G) = (L(G'))^R$

Since G' is right-linear, we have:

$L(G')$ is regular language

$(L(G'))^R$ is regular language (homework)

$L(G)$ is regular language

Proof - Part 2

Now we will prove:

If L is a regular language then there is a regular grammar G with $L(G) = L$

Proof outline:

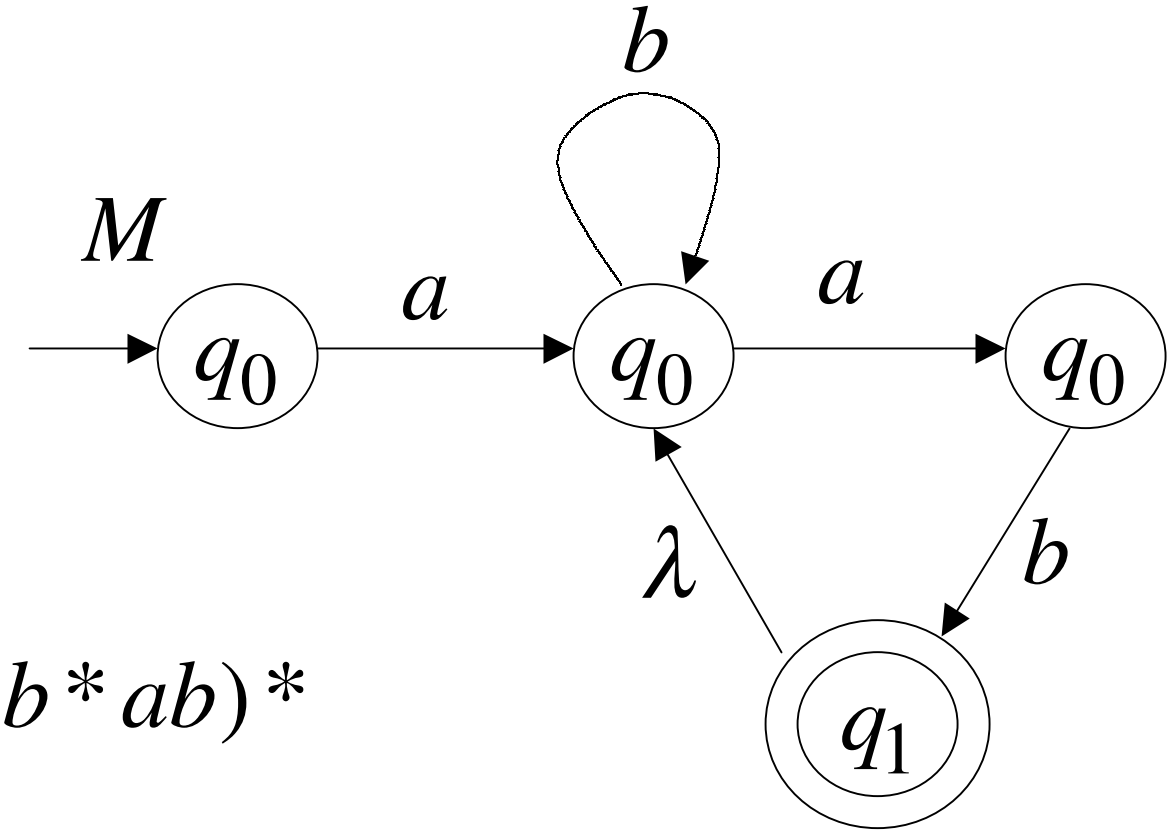
we will take an NFA for L

and convert it to a regular grammar

Since L is regular

There is an NFA M such that $L = L(M)$

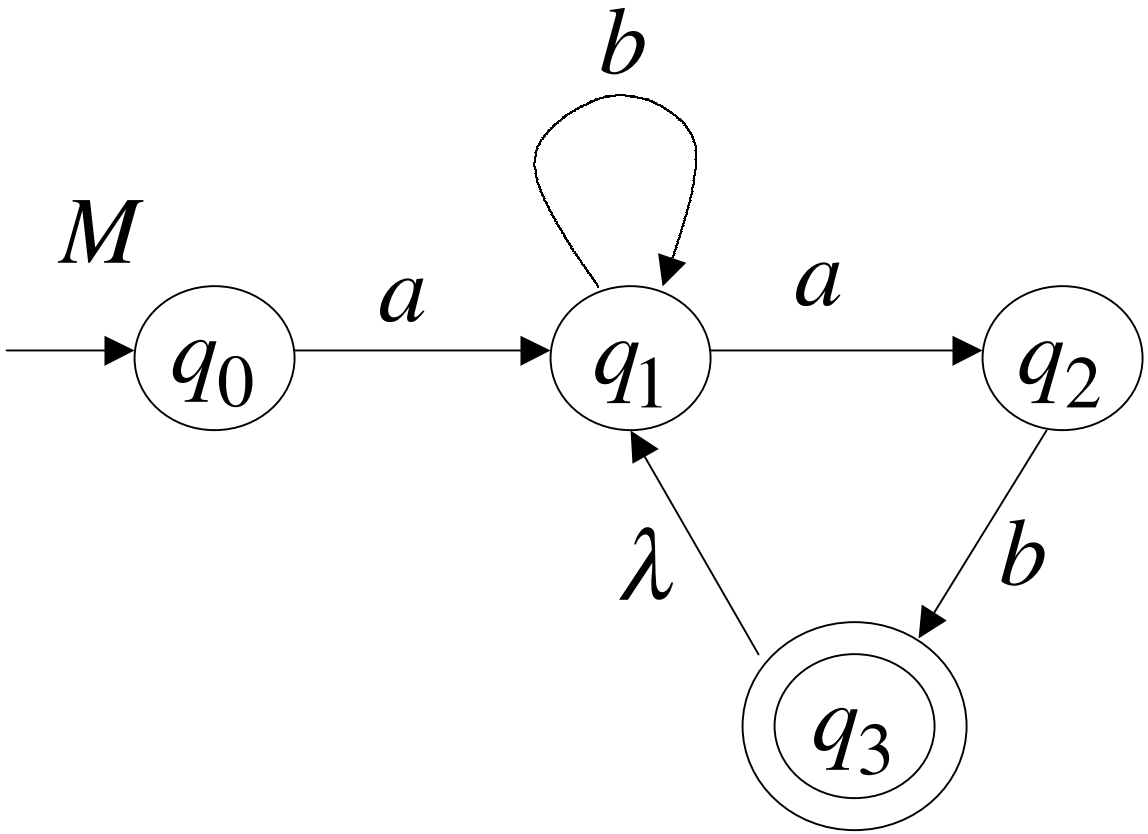
Example:



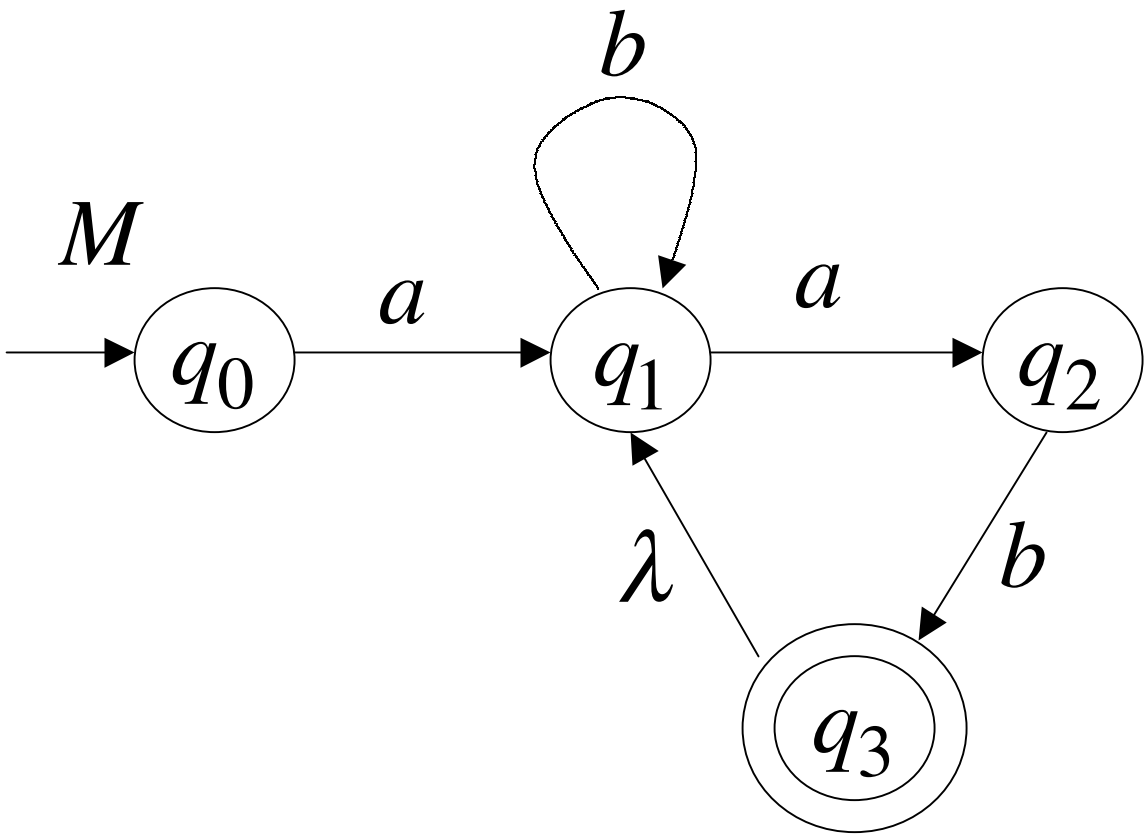
$$L = ab^*ab(b^*ab)^*$$

$$L = L(M)$$

Convert M to a right-linear grammar



$q_0 \to aq_1$



$q_0 \rightarrow aq_1$

$q_1 \rightarrow bq_1$

G

$q_0 \to aq_1$

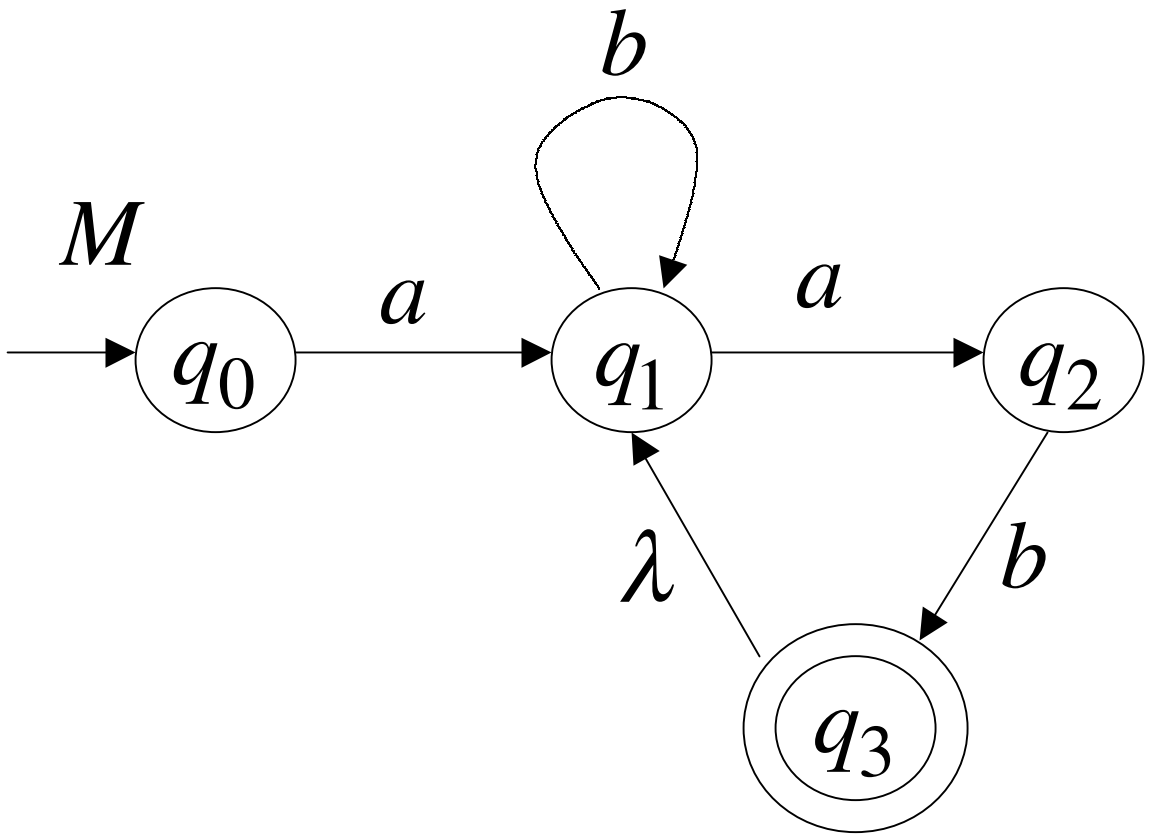
$q_1 \to bq_1$

$q_1 \to aq_2$

$q_2 \to bq_3$

$q_2 \to b$

$q_2 \to q_1$



We can generalize this process for any regular language L :

For any regular language L
we obtain an right-linear grammar G
with $L = L(G)$

Since G is right-linear grammar

G is also a regular grammar

with $L = L(G)$