Kellerautomaten

Eduard Mehofer

Fakultät für Informatik

Währinger Straße 29

Universität Wien

Kellerautomat (engl. Pushdown Automaton)

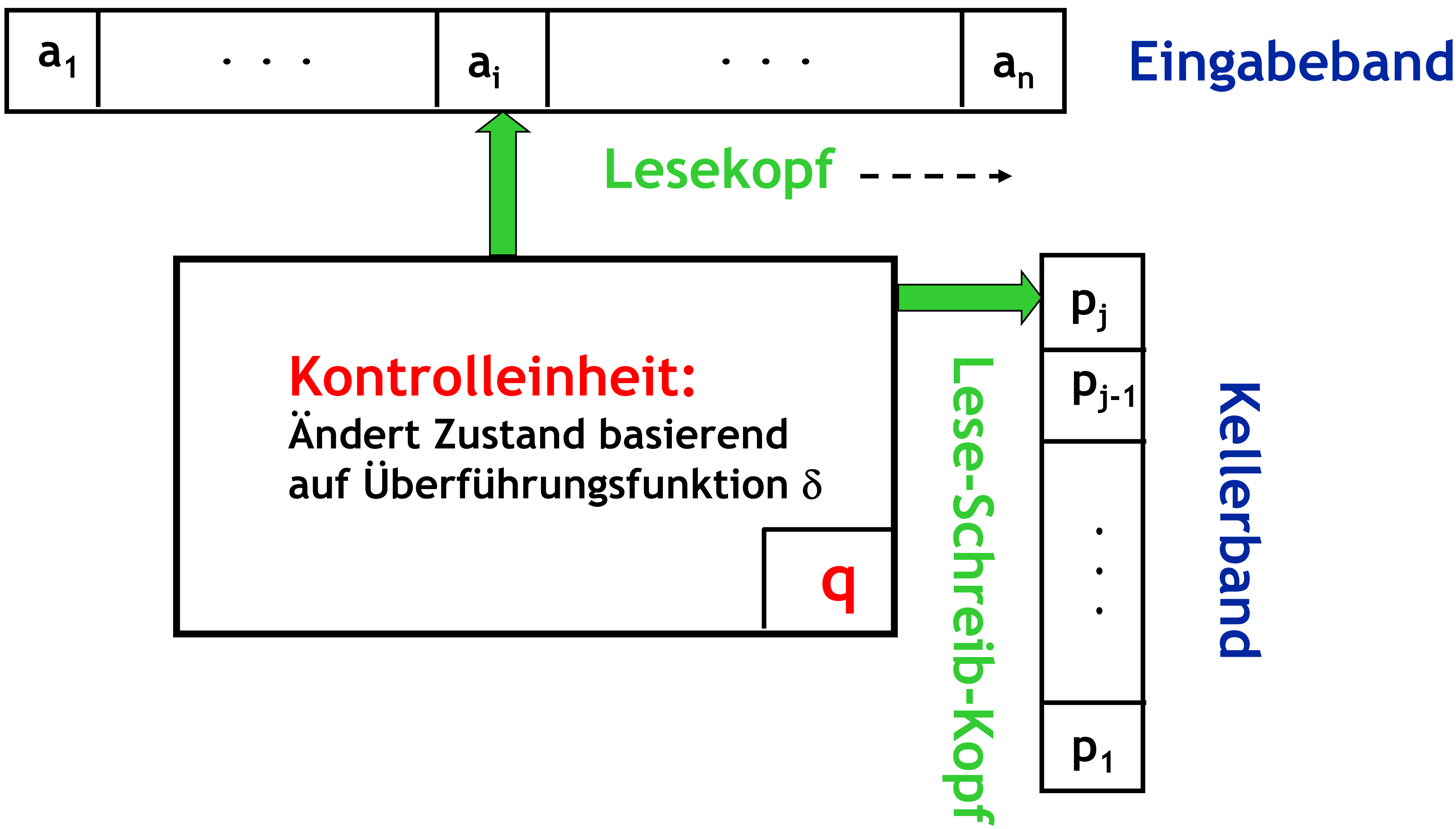
Ein Kellerautomat ist eine Erweiterung des ε-NEA um einen Keller.

Keller (Kellerspeicher, Stapelspeicher), *veraltend*: Speicher bei dem nur das oberste Symbol gelesen oder verändert werden darf, heute üblicherweise mit Stack bezeichnet (Zugriff mit push- und pop-Operationen).

Ein Kellerautomat besteht aus folgenden Komponenten:

- Eingabeband mit Lesekopf, der die Eingabesymbole von links nach rechts liest
- Kellerband mit Lese-Schreib-Kopf, der sich über dem obersten Kellersymbol befindet
- Kontrolleinheit mit endlich vielen Zuständen

Graphische Darstellung der Komponenten



Arbeitsweise eines Kellerautomaten (1)

Ein Kellerautomat startet in einem vorgegebenen Anfangszustand (q_0), hat den Lesekopf über dem ersten Symbol des Eingabewortes und hat den Lese-Schreib-Kopf über dem Startsymbol (Z_0) des ansonsten leeren Kellerbandes.

Die Aktionen eines Kellerautomaten werden durch die Überführungsfunktion δ definiert. δ wird durch eine Regelmenge beschrieben, die angibt, wie man von einer aktuellen Konfiguration des Kellerautomaten in eine Nachfolgekonfiguration gelangt.

Der Inhalt des Kellerbandes ist ein Wort, wobei von links nach rechts geschrieben das oberste Symbol des Kellerbandes am weitesten links steht, i.e. $p_j \, p_{j-1} \ldots p_1$.

Arbeitsweise eines Kellerautomaten (2)

In Abhängigkeit vom aktuellen Zustand und den beiden gelesenen Symbolen werden folgende Aktionen durchgeführt:

1. Der Lesekopf rückt um ein Symbol nach rechts oder bleibt über dem aktuellen Symbol (ε-Übergang).
2. Das oberste Kellersymbol wird gelöscht.
3. Ein Wort über dem Kelleralphabet oder ε wird auf das Kellerband geschrieben und der Lese-Schreib-Kopf wird auf das neue oberste Symbol des Kellerbandes gesetzt.
   - schreiben von ε: bewirkt eine pop-Operation (oberstes Symbol wird gelöscht)
   - schreiben vom gelöschten obersten Kellersymbol: Keller bleibt unverändert
   - schreiben von anderen Symbolen: Keller wird geändert
4. Der Kellerautomat geht in einen neuen Zustand über.

Kellerautomat KA

Definition: Ein (nichtdeterministischer) Kellerautomat ist ein 7-Tupel

$$KA = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$$

wobei

- **Q: Menge der Zustände (endlich)**
- **Σ: Eingabealphabet (endlich)**
- **Γ: Kelleralphabet (endlich)**
- **$q_0 \in Q$: Startzustand/Anfangszustand**
- **$Z_0 \in \Gamma$: Startsymbol/Anfangssymbol auf dem Kellerband**
- **$F \subseteq Q$: Menge der Endzustände (endlich)**
- **$\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \rightarrow$ Menge aus $P(Q \times \Gamma^*)$: Überführungsfunktion**
  (**P** … *Potenzmenge*)

<u>Man beachte</u>: Auf Stack darf ε (Γ^*) "geschrieben" werden (hat Semantik des Löschens), gelesen wird ε jedoch vom Stack nie (Γ), oder eine beliebige Zeichenkette über Γ.

Interpretation von δ

1. $\delta(q, a, Z) = \{ (q_1, \gamma_1), \ldots , (q_n, \gamma_n) \}$ mit
   $q_i \in Q$, $a \in \Sigma$, $Z \in \Gamma$, $\gamma_i \in \Gamma^*$

   Der Kellerautomat wählt nichtdeterministisch ein Paar (q_i, γ_i), aus, geht in den Zustand q_i über, löscht Z vom Kellerband, schreibt auf das Kellerband γ_i, rückt den Lesekopf um eins nach rechts und setzt den Lese-Schreib-Kopf über das erste Symbol von γ_i.

2. $\delta(q, \varepsilon, Z) = \{ (q_1, \gamma_1), \ldots , (q_n, \gamma_n) \}$

   Der Lesekopf liest nichts und rückt daher auch nicht nach rechts, sonst bleibt alles ident zu 1, d.h. ein Zustandswechsel ohne ein Eingabesymbol zu konsumieren.
   <u>Achtung</u>: ε bedeutet <u>NICHT</u> , dass das Eingabeband leer ist!

<u>Man beachte</u>: $\delta(q, a, \varepsilon)$ mit ε in dritter Position ist nicht erlaubt!

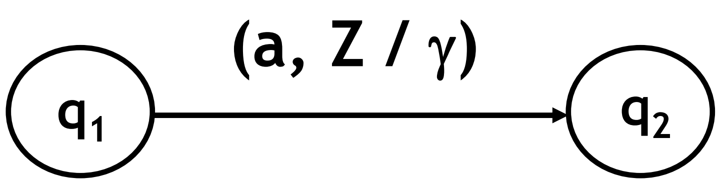
Übergangsdiagramm eines KA

Definition: Das Übergangsdiagramm für einen KA ist ein gerichteter Graph, der wie folgt definiert ist:

- Die Knoten des Graphen entsprechen den Zuständen des KA.
- Eine mit **$(a, Z / \gamma)$** bezeichnete Kante von Zustand **q_1** nach Zustand **q_2** beschreibt **$(q_2,\gamma) \in \delta(q_1, a, Z)$**.

Der Anfangszustand wird durch einen mit "**Start**" markierten Pfeil angegeben und die Endzustände werden mit einem **doppelten Kreis** markiert.

Bsp.:



Konfiguration und Überführungsrelation (1)

Eine **Konfiguration** (Situation, Momentaufnahme) eines Kellerautomaten KA = $(Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ ist ein Tripel $(q, w, \gamma) \in Q \times \Sigma^* \times \Gamma^*$ mit

- q aktueller Zustand,
- w zu verarbeitendes Restwort rechts vom Lesekopf,
- γ aktueller Kellerinhalt.

Dieses Tripel bedeutet, dass sich der Kellerautomat im Zustand q befindet, dass das Wort rechts vom Lesekopf gleich w ist und dass γ am Kellerband steht.

Die **Überführungsrelation** $\vdash$ führt eine Konfiguration (q_1, w_1, γ_1) in eine Nachfolgekonfiguration (q_2, w_2, γ_2) über, i.e.

$$(q_1, w_1, \gamma_1) \vdash (q_2, w_2, \gamma_2)$$

Konfiguration und Überführungsrelation (2)

Die **Überführungsrelation** $(Q \times \Sigma^* \times K^*) \times (Q \times \Sigma^* \times K^*)$ ist wie folgt definiert:

- $(q_1, aw, Z\gamma) \vdash (q_2, w, \gamma_1\gamma)$ genau dann, wenn $(q_2,\gamma_1) \in \delta(q_1, a, Z)$

  mit $q_1, q_2 \in Q$, $a \in \Sigma \cup \{\varepsilon\}$, $w \in \Sigma^*$, $Z \in \Gamma$ und $\gamma, \gamma_1 \in \Gamma^*$

Die Relation $\vdash^{*}_{KA}$ ist die **reflexive und transitive Hülle** von $\vdash_{KA}$.

Akzeptierte Sprache

Definition: Sei KA = $(Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ ein Kellerautomat. Dann ist

$$\mathbf{L(KA)} = \{ w \in \Sigma^* \mid (q_0, w, Z_0) \vdash^*_{KA} (q, \varepsilon, \gamma) \text{ mit } q \in F \text{ und } \gamma \in \Gamma^* \}$$

die von KA durch **Endzustand** akzeptierte Sprache.

Man beachte:

a) Ein Kellerautomat akzeptiert ein Wort nur, wenn es **zur Gänze** gelesen wurde und man sich in einem Endzustand befindet (analog zu einem endlichen Automaten).

b) Ein leeres Kellerband ist in dieser Definition nicht gefordert.

Anm.: Dieser Akzeptanz-Ansatz wird in der VO und UE verwendet und wird bei den Beispielen durch die Angabe "durch Endzustand" verdeutlicht.

Alternative Sprachakzeptanz-Regeln

Sei KA = (Q, Σ, Γ, δ, q_0, Z_0, F) ein Kellerautomat. Dann ist

1) Sprache L(KA), die von KA durch Endzustand akzeptiert wird, ist

   $L(KA) = \{ w \in \Sigma^* \mid (q_0, w, Z_0) \vdash_{KA}^{*} (q, \varepsilon, \gamma) \text{ mit } q \in F \text{ und } \gamma \in \Gamma^* \}$

2) Sprache N(KA), die von KA durch leeres Kellerband akzeptiert wird, ist

   $N(KA) = \{ w \in \Sigma^* \mid (q_0, w, Z_0) \vdash_{KA}^{*} (q, \varepsilon, \varepsilon) \text{ mit } q \in Q \}$

3) Sprache LN(KA), die von KA durch Endzustand und leeres Kellerband akzeptiert wird, ist

   $LN(KA) = \{ w \in \Sigma^* \mid (q_0, w, Z_0) \vdash_{KA}^{*} (q, \varepsilon, \varepsilon) \text{ mit } q \in F \}$

Sprachen und Kellerautomaten

Gegeben sei eine kontextfreie Grammatik G mit L = L(G). Dann gilt folgendes:

1) Es gibt einen Kellerautomaten KA mit L = L(KA). (*Endzustand*)

2) Es gibt einen Kellerautomaten KA mit L = N(KA). (*leeres Kellerband*)

3) Es gibt einen Kellerautomaten KA mit L = LN(KA). (*Endzustand plus leeres Kellerband*)

D.h. eine kontextfreie Sprache kann je nach persönlicher Vorliebe unter allen 3 Akzeptanzregeln akzeptiert werden. Wie bereits erwähnt wählen wir in dieser Lehrveranstaltung in Analogie zu den endlichen Automaten die <u>Akzeptanz durch Endzustand</u> (Variante 1).

Deterministischer KA

Bei einem det. KA gibt es für jede Konfiguration (q, aw, Zγ) höchstens **eine** Folgekonfiguration.

Definition: Ein Kellerautomat KA = (Q, Σ, Γ, δ, q_0, Z_0, F) ist **deterministisch**, falls für alle q $\in$ Q, a $\in$ Σ, Z $\in$ Γ gilt, dass

1. $|\ \delta(q, a, Z)\ | \leq 1$

2. $|\ \delta(q, \varepsilon, Z)\ | \leq 1$

3. falls $\delta(q, \varepsilon, Z) \neq \varnothing$, dann $\delta(q, a, Z) = \varnothing$ für alle a $\in \Sigma$

   d.h. in Zustand q mit Kellersymbol Z hat man gar keine andere Möglichkeit als im Übergangsdiagramm die ε-Kante zu wählen.

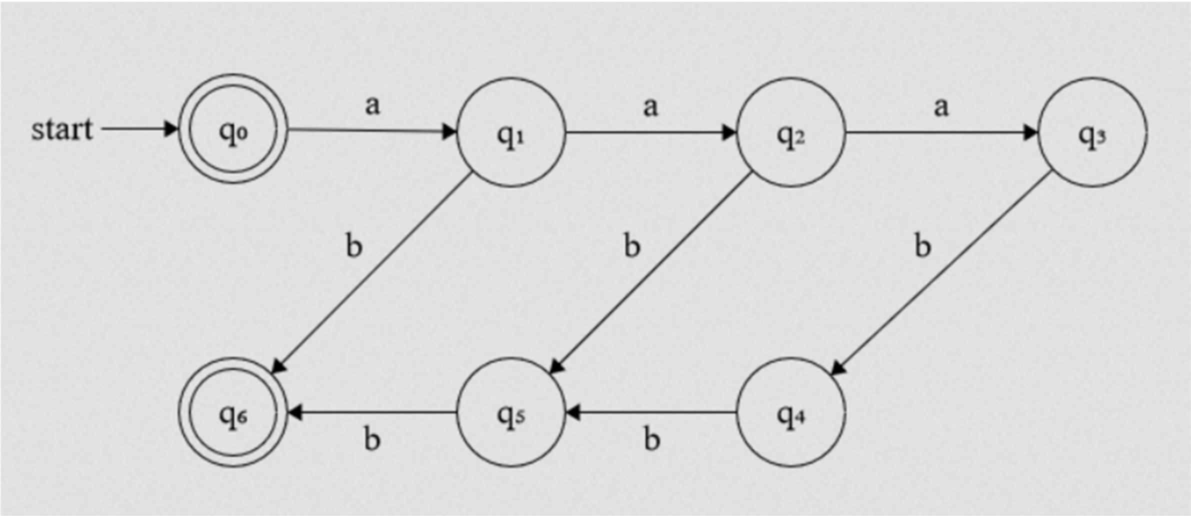
Beispiele mit KA

Kellerautomaten sind gut geeignet für Sprachen, bei denen eine Akzeptanz das Abzählen oder Merken von Symbolen erfordert. Beispiele sind:

- $L = \{a^n b^n |\, n \geq 0\}$
- $L = \{a^n b^{2n} |\, n \geq 0\}$
- $L = \{a^n b^m c^n |\, m, n \geq 0\}$
- wohlgeformte Klammerausdrücke, i.e. öffnende und schließende Klammern entsprechen sich

Obige Sprachen können mit einem endlichen Automaten nur für Sonderfälle akzeptiert werden, z.B. für

- $L = \{a^n b^n |\, n \geq 0 \wedge n \leq 3\}$
  <u>endlicher Automat</u>
  A=($\{q_0,q_1,q_2,q_3,q_4,q_5,q_6\}$,
  $\{a,b\},\delta,q_0,\{q_0,q_6\}$)

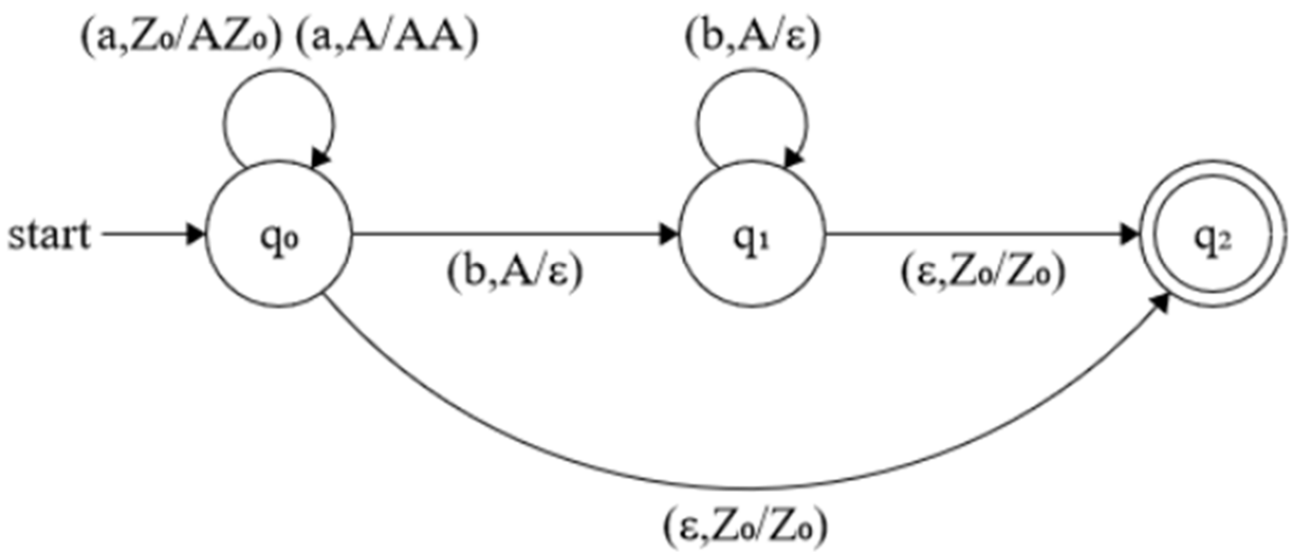


Beispiel 1a: Kellerautomat für

- $L = \{a^n b^n | \, n \geq 0\}$

KA = (Q, Σ, Γ, δ, q_0, Z_0, F)

$\Sigma = \{a,b\}$, $Q = \{q_0,q_1,q_2\}$, $\Gamma = \{A,Z_0\}$, $F = \{q_2\}$



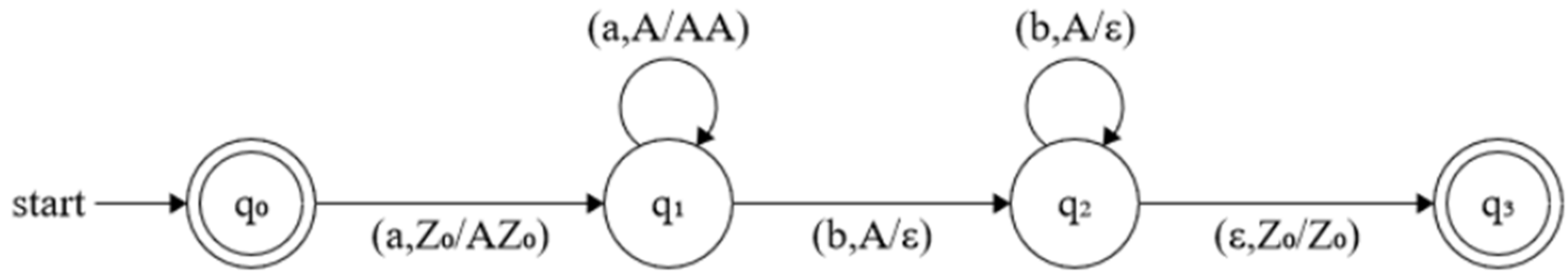
Ist der Kellerautomat deterministisch? Antwort: **NEIN**

Beispiel 1b: deterministischer Kellerautomat für

- $L = \{a^n b^n | \, n \geq 0\}$

KA = (Q, Σ, Γ, δ, q_0, Z_0, F)

Σ = {a,b}, Q = {q_0,q_1,q_2,q_3}, Γ = {A,Z_0}, F = {q_0,q_3}



Eingabewort aaabbb wird durch Endzustand akzeptiert:

$(q_0,aaabbb,Z_0) \vdash (q_1,aabbb, AZ_0) \vdash (q_1,abbb,AAZ_0) \vdash (q_1,bbb,AAAZ_0) \vdash$
$(q_2,bb,AAZ_0) \vdash (q_2,b,AZ_0) \vdash (q_2,\varepsilon,Z_0) \vdash (q_3,\varepsilon,Z_0)$

Beispiel 2:

Gesucht ist ein Kellerautomat für die Sprache $L = \{ v v^R \mid v \in \{0, 1\}^* \}$, wobei v^R das gespiegelte Wort von v bezeichnet, i.e. Palindrome gerader Länge über {0,1}, der <u>durch Endzustand akzeptiert</u>.

Grammatik $G=(\{S\}, \{0,1\}, \{ S \rightarrow 0S0 \mid 1S1 \mid \varepsilon \}, S)$ mit $L(G)=L$.

Für diese Sprache gibt es **keinen** deterministischen Kellerautomaten!

Strategie: v wird eingelesen und eine Kopie im Keller gespeichert, die Mitte muss geraten werden, und v^R wird mit den Symbolen im Keller verglichen.

$KA = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$

$\Sigma = \{0, 1\}$, $Q = \{q_0, q_1, q_2\}$, $\Gamma = \{0, 1, Z_0\}$, $F = \{q_2\}$

Überführungsfunktion **δ** :

Bearbeitung des ersten Symbols von v – wird in den Keller gegeben:

1. $\delta(q_0, 0, Z_0) = \{(q_0, 0Z_0)\}$
2. $\delta(q_0, 1, Z_0) = \{(q_0, 1Z_0)\}$

Beispiel 2 (Fortsetzung):

Alle weiteren Symbole von v werden im Keller abgespeichert:

3. $\delta(q_0, 0, 0) = \{(q_0, 00)\}$
4. $\delta(q_0, 0, 1) = \{(q_0, 01)\}$
5. $\delta(q_0, 1, 0) = \{(q_0, 10)\}$
6. $\delta(q_0, 1, 1) = \{(q_0, 11)\}$

Ist die Mitte des Eingabewortes erreicht, wechselt man nichtdeterministisch mit einem ε-Übergang in den Zustand q_1, wobei der Keller nicht verändert wird. Die Mitte wird erraten:

7. $\delta(q_0, \varepsilon, Z_0) = \{(q_1, Z_0)\}$
8. $\delta(q_0, \varepsilon, 0) = \{(q_1, 0)\}$
9. $\delta(q_0, \varepsilon, 1) = \{(q_1, 1)\}$

Beispiel 2 (Fortsetzung):

Im Zustand q_1 gilt die Annahme, dass die Mitte des Wortes erreicht wurde und die Übereinstimmung von v mit v^R geprüft werden muss. Die Eingabesymbole und Kellersymbole müssen übereinstimmen und der Keller wird abgebaut, d.h. bei Übereinstimmung wird eine pop-Operation durchgeführt:

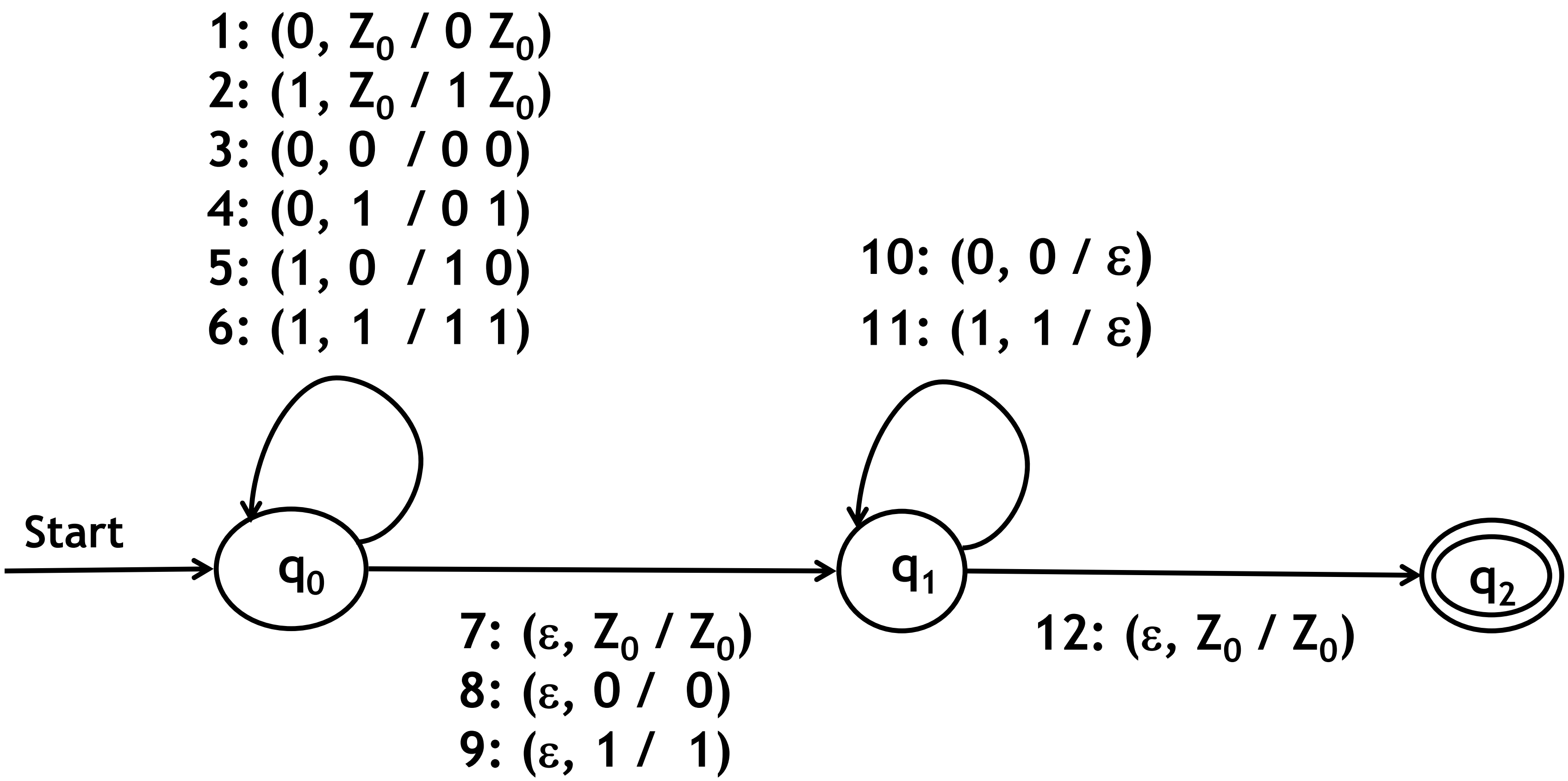
10. $\delta(q_1, 0, 0) = \{(q_1, \varepsilon)\}$
11. $\delta(q_1, 1, 1) = \{(q_1, \varepsilon)\}$

Stößt man auf das Startsymbol des Kellers, geht man in den Endzustand über:

12. $\delta(q_1, \varepsilon, Z_0) = \{(q_2, Z_0)\}$

Das Wort wird im Endzustand akzeptiert, wenn es zur Gänze gelesen wurde, sonst nicht. Das Kellerband ist in obiger Lösung nicht leer. Wird (12) durch $\delta(q_1, \varepsilon, Z_0) = \{(q_2, \varepsilon)\}$ ersetzt um das Startsymbol des Kellers zu löschen, ist der Keller bei Akzeptierung leer.

Beispiel 2 (Fortsetzung): Übergangsdiagramm



Beispiel 2 (Fortsetzung): Analyse

Nichtdeterministische Analyse des Eingabewortes **1111**:

a) $(q_0, 1111, Z_0) \vdash (q_0, 111, 1Z_0) \vdash (q_0, 11, 11Z_0) \vdash$
$(q_1, 11, 11Z_0) \vdash (q_1, 1, 1Z_0) \vdash (q_1, \varepsilon, Z_0) \vdash (q_2, \varepsilon, Z_0)$

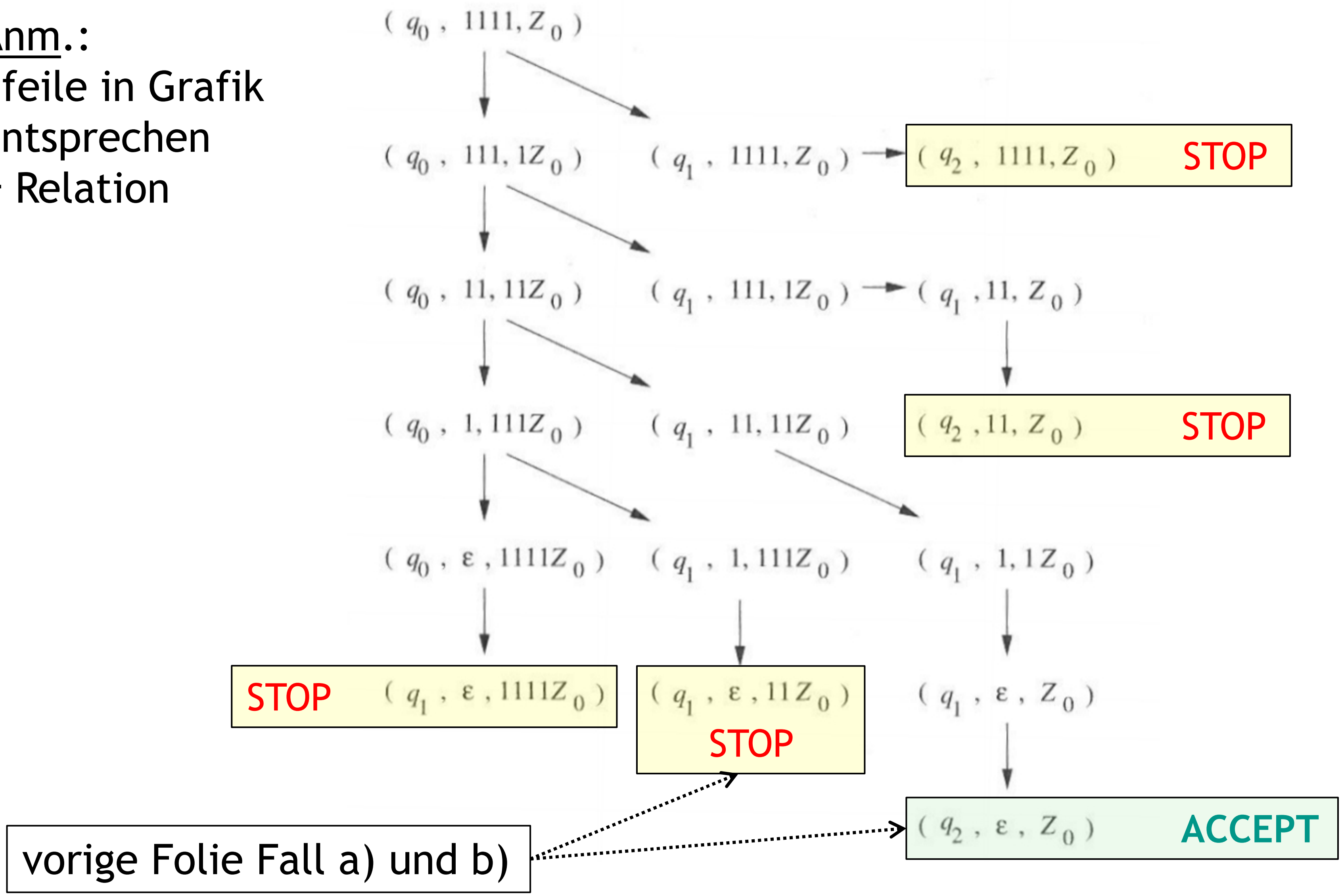
Der Kellerautomat akzeptiert das Eingabewort **1111** durch Endzustand.

b) $(q_0, 1111, Z_0) \vdash (q_0, 111, 1Z_0) \vdash (q_0, 11, 11Z_0) \vdash$
$(q_0, 1, 111Z_0) \vdash (q_1, 1, 111Z_0) \vdash (q_1, \varepsilon, 11Z_0)$

Der Kellerautomat erreicht bei Eingabewort **1111** keinen Endzustand.

Beispiel 2 (Fortsetzung): Analysemöglichkeiten

<u>Anm.</u>:
Pfeile in Grafik
entsprechen
⊢ Relation



Beispiel 3a:

Gesucht ist ein deterministischer Kellerautomat für die Sprache L = { vcv^R | $v \in \{0, 1\}^*$}, wobei v^R das gespiegelte Wort von v ist, i.e. Palindrome ungerader Länge über {0,1} mit Mittemarkierung c, der <u>durch Endzustand akzeptiert</u>.
Grammatik G=({S}, {0,1,c}, { S $\rightarrow$ 0S0|1S1|c }, S) mit L(G)=L.
Strategie ident zu Beispiel 1.

KA = $(Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$
$\Sigma = \{0, 1, c\}$, $Q = \{q_0, q_1, q_2\}$, $\Gamma = \{0,1,Z_0\}$ (auch möglich: $\Gamma = \{X,Y,Z_0\}$), $F = \{q_2\}$

Überführungsfunktion δ:
Bearbeitung des ersten Symbols von v:

1. $\delta(q_0, 0, Z_0) = \{(q_0, 0Z_0)\}$
2. $\delta(q_0, 1, Z_0) = \{(q_0, 1Z_0)\}$

Bearbeitung der folgenden Symbole von v:

3. $\delta(q_0, 0, 0) = \{(q_0, 00)\}$
4. $\delta(q_0, 0, 1) = \{(q_0, 01)\}$
5. $\delta(q_0, 1, 0) = \{(q_0, 10)\}$
6. $\delta(q_0, 1, 1) = \{(q_0, 11)\}$

Beispiel 3a (Fortsetzung)

Bei Auftreten von c geht der Automat in den Zustand q_1 über, der Keller bleibt unverändert:

7. $\delta(q_0, c, Z_0) = \{(q_1, Z_0)\}$
8. $\delta(q_0, c, 0) = \{(q_1, 0)\}$
9. $\delta(q_0, c, 1) = \{(q_1, 1)\}$

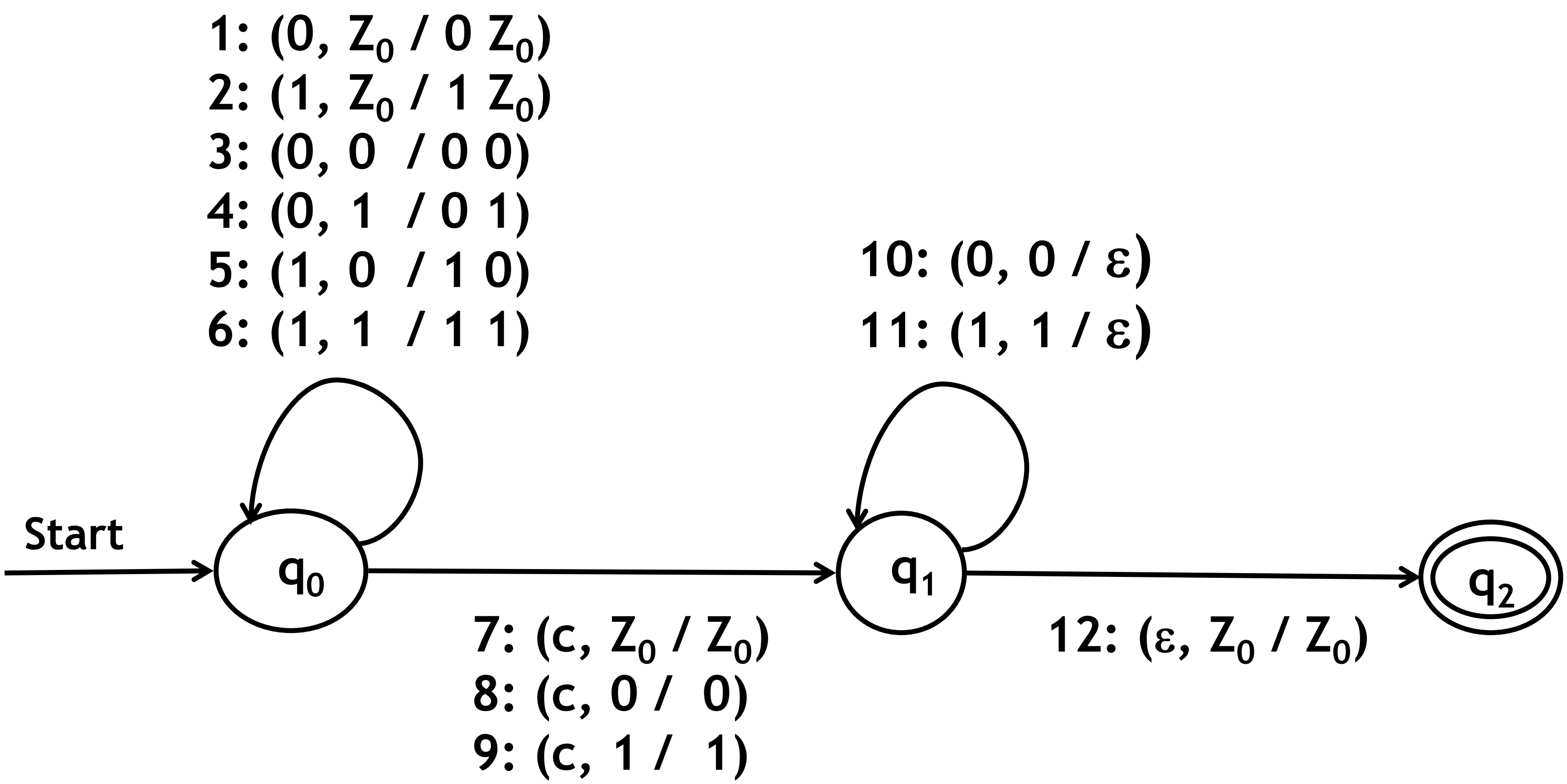
Im Zustand q_1 wird die Spiegelung überprüft, d.h. Eingabesymbole und Kellersymbole müssen übereinstimmen:

10. $\delta(q_1, 0, 0) = \{(q_1, \varepsilon)\}$ (es werden die obersten Kellersymbole gelöscht)
11. $\delta(q_1, 1, 1) = \{(q_1, \varepsilon)\}$

Bei Startsymbol des Kellers geht der Automat in den Endzustand über:

12. $\delta(q_1, \varepsilon, Z_0) = \{(q_2, Z_0)\}$

Beispiel 3a (Fortsetzung): Übergangsdiagramm



Beispiel 3a (Fortsetzung): Analyse

Analyse des Eingabewortes **01c100**:

$(q_0, 01c100, Z_0) \vdash (q_0, 1c100, 0Z_0) \vdash (q_0, c100, 10Z_0) \vdash$

$(q_1, 100, 10Z_0) \vdash (q_1, 00, 0Z_0) \vdash (q_1, 0, Z_0) \vdash$

$(q_2, 0, Z_0)$

Der Kellerautomat akzeptiert das Eingabewort **01c100** nicht, da das Eingabewort nicht zur Gänze abgearbeitet wurde.

Anmerkungen zu Beispiel 3a:

Die Regeln 10 und 11

10. $\delta(q_1, 0, 0) = \{(q_1, \varepsilon)\}$
11. $\delta(q_1, 1, 1) = \{(q_1, \varepsilon)\}$

erkennen nicht die letzte Ziffer des Wortes, d.h. man befindet sich in Zustand q_1 und das gesamte Wort wurde bereits gelesen. Die einzige Möglichkeit, das Wort noch zu akzeptieren, ist ein ε-Übergang mit Regel 12

12. $\delta(q_1, \varepsilon, Z_0) = \{(q_2, Z_0)\}$

Man beachte aber, dass der Kellerautomat deterministisch ist, da man in q_1 bei Kellersymbol Z_0 nur den ε-Übergang wählen kann.

In Beispiel 3b wollen wir eine Lösung entwickeln, die keinen ε-Übergang benötigt.

Beispiel 3b: Alternative Lösung zu 3a

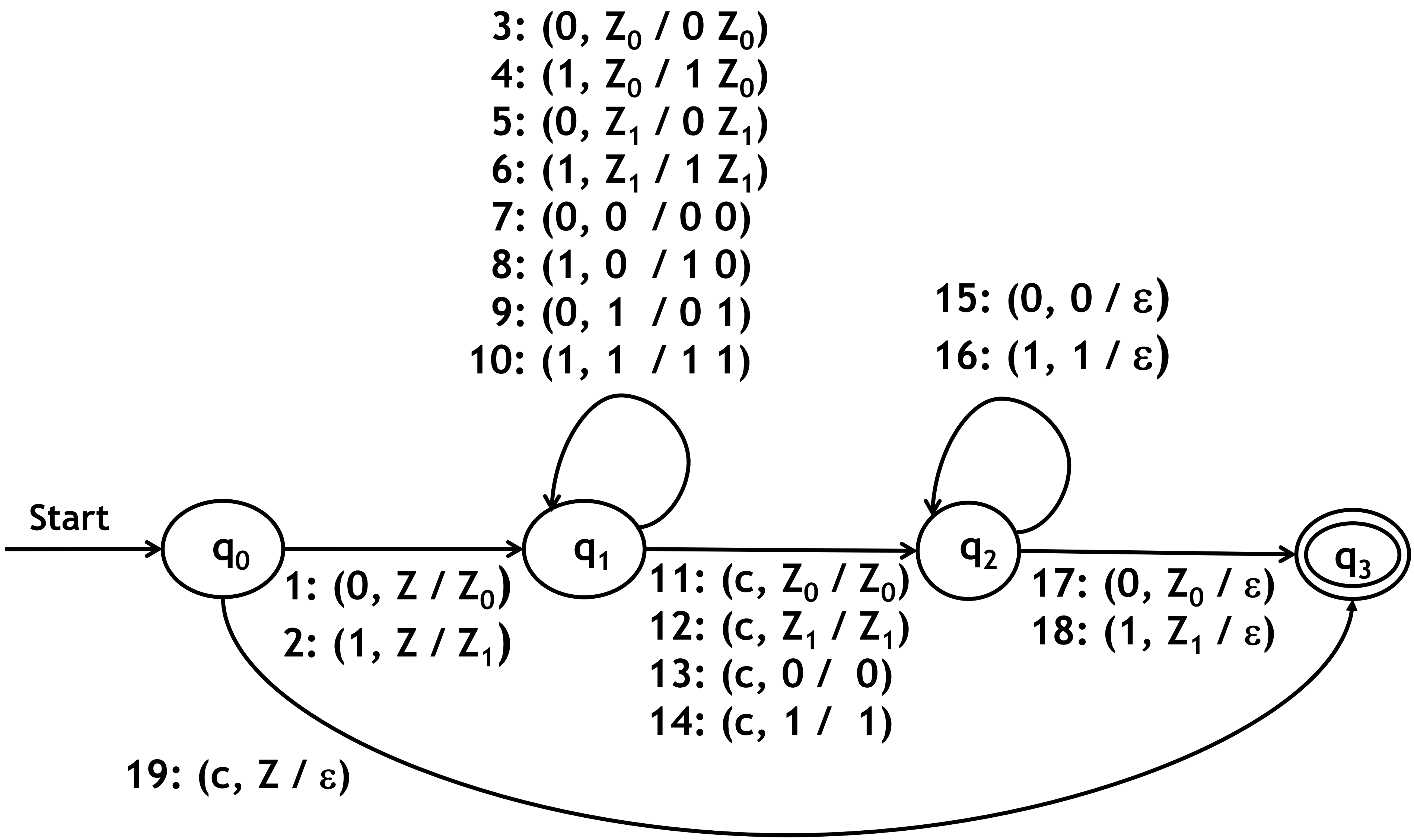
Nun wollen wir einen alternativen Kellerautomaten entwickeln, der anders als in 3a keinen ε-Übergang benötigt. Wie zu 3a angemerkt, muss bei dieser Lösung die letzte Ziffer eines korrekten Wortes erkannt werden, damit ich ohne ε in den Endzustand wechseln kann.

Um dies zu erreichen geben wir für die erste Ziffer des Wortes das Kellersymbol Z_0 für 0 bzw. Z_1 für 1 in den Keller. Sobald wir während der Abarbeitung eines Wortes auf diese Kellersymbole stoßen, müssen wir in den Endzustand wechseln.

$KA' = (Q, \Sigma, \Gamma, \delta, q_0, Z, F)$

$\Sigma = \{0, 1, c\}$, $Q = \{q_0, q_1, q_2, q_3\}$, $\Gamma = \{0, 1, Z_0, Z_1, Z\}$, $F = \{q_3\}$

Beispiel 3b (Fortsetzung): Übergangsdiagramm



Beispiel 3b (Fortsetzung):

Die Regeln 17 und 18 erkennen die letzte Ziffer des Analysewortes, sofern es sich um ein korrektes Wort aus L handelt, und bewirken einen Wechsel in den Endzustand. Sind noch Symbole am Eingabeband, die noch nicht gelesen wurden, wird das Wort lt. Definition Kellerautomat trotz Endzustand **nicht** akzeptiert.

Man beachte, dass die Lösung 3b Worte aus L durch Endzustand und leeres Kellerband akzeptiert.

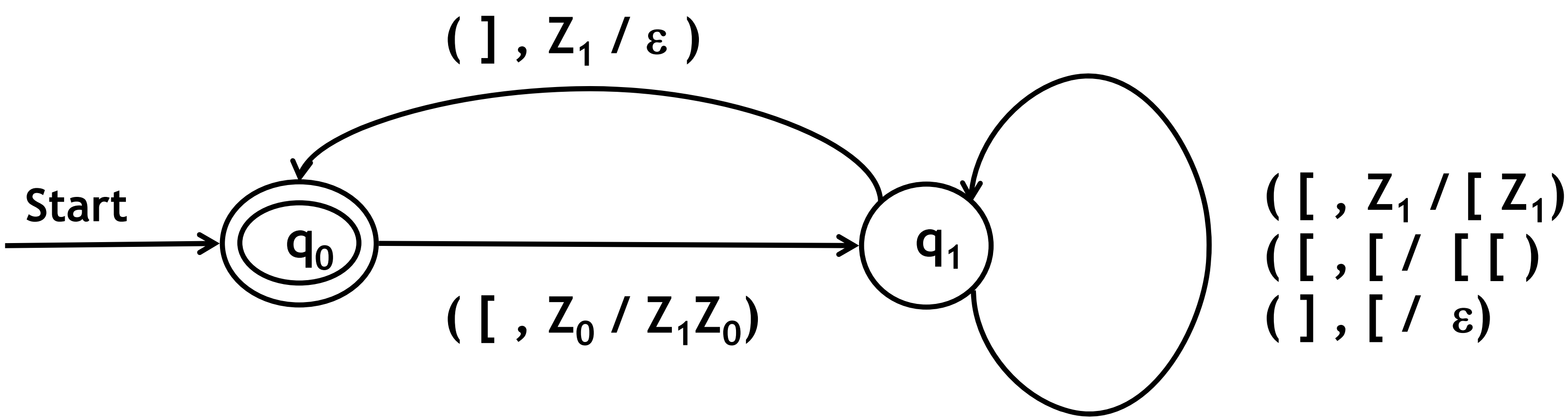
Beispiel 4:

Gesucht ist ein deterministischer Kellerautomat, der wohlgeformte Klammer-ausdrücke durch Endzustand akzeptiert, die durch die KFG G = ({S}, {[,]}, { S $\rightarrow$ [S]S, S$\rightarrow\varepsilon$ }, S) definiert ist.

KA = (Q, Σ, Γ, δ, q_0, Z_0, F)

Σ = { [,] }, Q = {q_0, q_1}, Γ = { [, Z_1, Z_0}, F = {q_0}

Überführungsfunktion δ:



Abschließende Anmerkungen

Satz: Sei KA=(Q, Σ, Γ, δ, q_0, Z_0, F) ein Kellerautomat, dann existiert eine kontextfreie Grammatik G mit L(G)=L(KA).

Satz: Sei G=(N,Σ,P,S) eine kontextfreie Grammatik, dann existiert ein Kellerautomat KA mit L(KA)=L(G).

Man beachte: $L_{det\text{-}KA} \subset L_{nichtdet\text{-}KA}$

Nichtdeterministische Kellerautomaten sind mächtiger als deterministische Kellerautomaten:

- nichtdet. KA akzeptieren alle kontextfreie Sprachen
- det. KA akzeptieren alle regulären Sprachen, aber bloß eine Teilmenge der kontextfreien Sprachen
- det. KA sind für Parser von Compilern wichtig, d.h. die Syntax von Programmiersprachen wird i.a. so gewählt, dass sie von einem det. KA analysiert werden kann