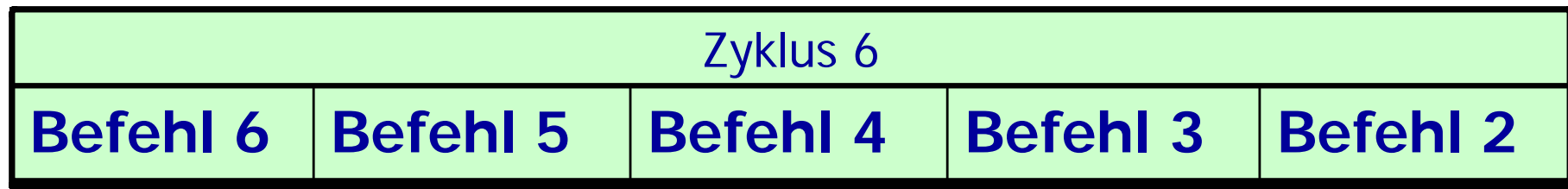
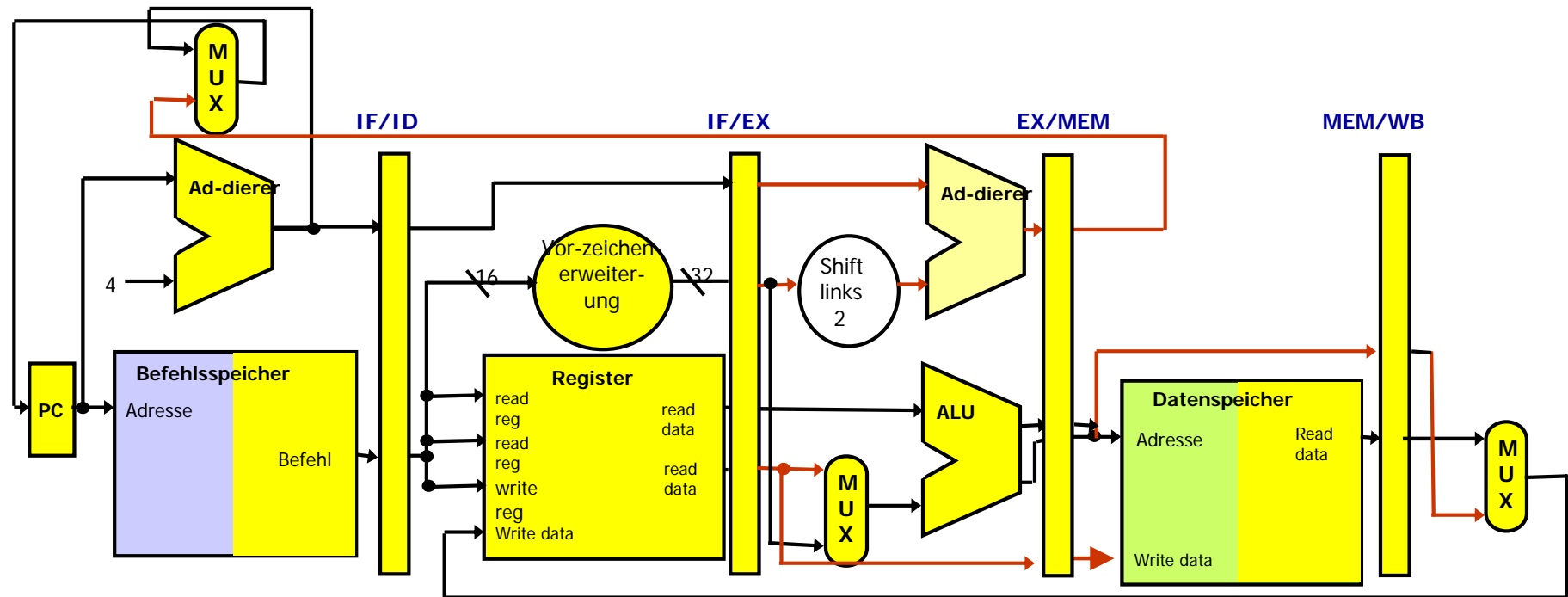


5.5 Pipeline Konflikte

- Bei einer gefüllten Pipeline wird pro Taktzyklus ein Befehl beendet.



5.5 Pipeline Konflikte

- Es gibt leider mehrere potentielle Probleme, die den Durchfluss durch die Pipeline hemmen bzw. verzögern. Man spricht von **Pipeline-Hemmnissen**
- Diese Verzögerungen entstehen durch **Daten-, Struktur- und Steuerflussabhängigkeiten**



Drei Arten von Pipeline-Konflikten

- **Datenkonflikte:** Treten auf, wenn ein Operand ist in der Pipeline (noch) nicht verfügbar.
 - Datenkonflikte werden durch Datenabhängigkeiten im Befehlsstrom erzeugt
- **Struktur- oder Ressourcenkonflikte:** Treten auf, wenn zwei Pipeline-Stufen dieselbe Ressource benötigen, auf diese aber nur einmal zugegriffen werden kann.
- **Steuerflusskonflikte** treten bei Programmsteuerbefehlen auf:
 - wenn in der Befehlsbereitstellungsphase die Zieladresse des als nächstes auszuführenden Befehls noch nicht berechnet ist bzw.
 - wenn im Falle eines bedingten Sprunges noch nicht klar ist, ob überhaupt gesprungen wird.



Beispiel: Datenabhängigkeiten

S_1 : **add** **r1, r2, 2** # **r1 := r2 + 2**

S_2 : **add** **r4, r1, r3** # **r4 := r1 + r3**

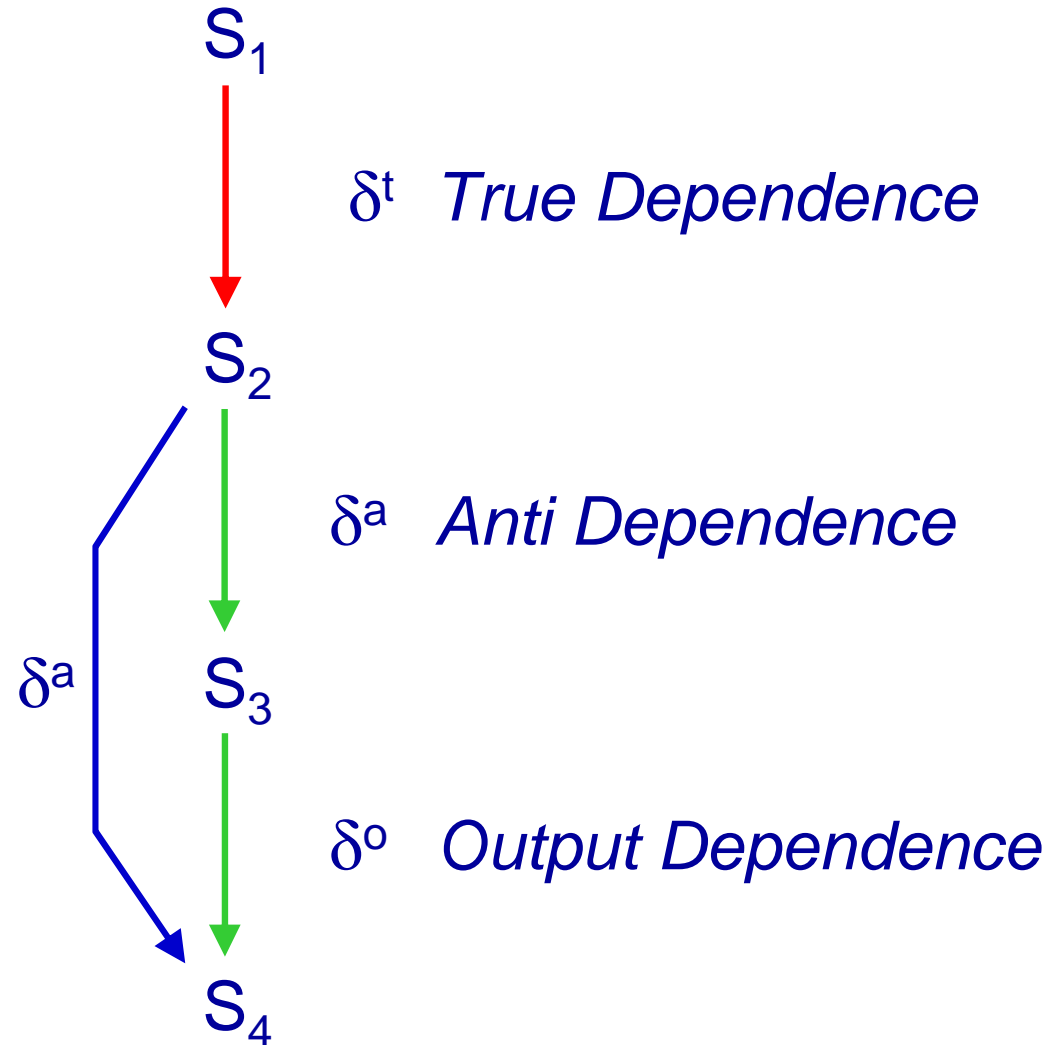
S_3 : **mul** **r3, r5, 3** # **r3 := r5 * 3**

S_4 : **mul** **r3, r6, 3** # **r3 := r6 * 3**



Beispiel: Datenabhängigkeiten

S_1 : **add** **r1, r2, 2**
 S_2 : **add** **r4, r1, r3**
 S_3 : **mul** **r3, r5, 3**
 S_4 : **mul** **r3, r6, 3**



5.5.1 Datenabhängigkeiten

Bemerkung:

Gegenabhängigkeiten und Ausgabeabhängigkeiten werden häufig auch **falsche Datenabhängigkeiten** oder entsprechend dem englischen Begriff „*Name Dependency*“ **Namensabhängigkeiten** genannt.



5.5.2 Datenkonflikte

- ❑ **Lese-nach-Schreibe-Konflikt** (read after write, RAW): Wird durch echte Abhängigkeit verursacht.
- ❑ **Schreibe-nach-Lese-Konflikt** (write after read, WAR): Wird durch Gegenabhängigkeit verursacht. Tritt dann auf, wenn in einer Pipeline die Schreibstufe der Lesestufe vorangeht.
- ❑ **Schreibe-nach-Schreibe-Konflikt** (write after write, WAW): Wird durch Ausgabeabhängigkeit verursacht. Tritt in Pipelines auf, die mehr als in einer Stufe schreiben, oder es erlauben, dass die Ausführung eines Befehls fortgesetzt werden darf, wenn ein vorhergehender Befehl angehalten worden ist.



WAR und WAW

Können WAR und WAW in der fünfstufigen DLX-Pipeline auftreten?

Nein, weil:

- Alle Befehle werden in 5 Stufen ausgeführt,
- Lesen aus Registern immer in der Stufe 2, und
- Schreiben in Register immer in der Stufe 5.

WAR und WAW treten bei "komplexeren" Pipelines auf



Beispiel

load r1,A



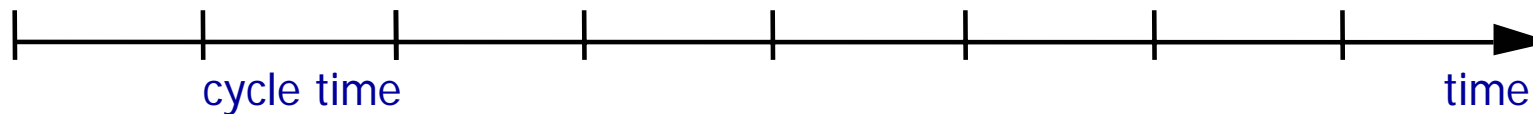
load r2,B



add r2,r1,r2



mul r1,r2,r1



5.5.2 Lösungen für Datenkonflikte

□ Software-Lösung

➤ Aufgabe des Compilers:

- Erkennen von Datenkonflikten
- Einfügen von Leeroperationen (`noop`) nach jedem Befehl, der einen Konflikt verursacht oder verursachen kann.

➤ Statische Befehlsanordnung:

- Instruction Scheduling, Pipeline Scheduling
- Umordnen der Befehle des Programms (Code-Optimierung), um Leeroperationen zu eliminieren.

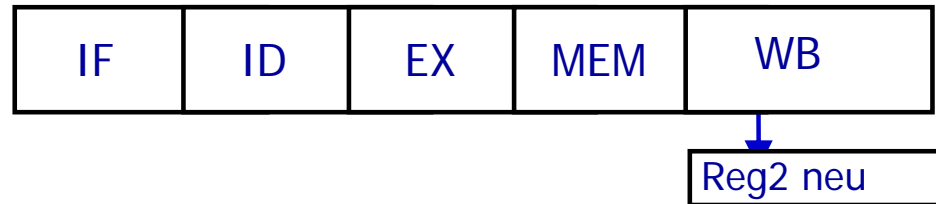


5.5.2 Lösungen für Datenkonflikte

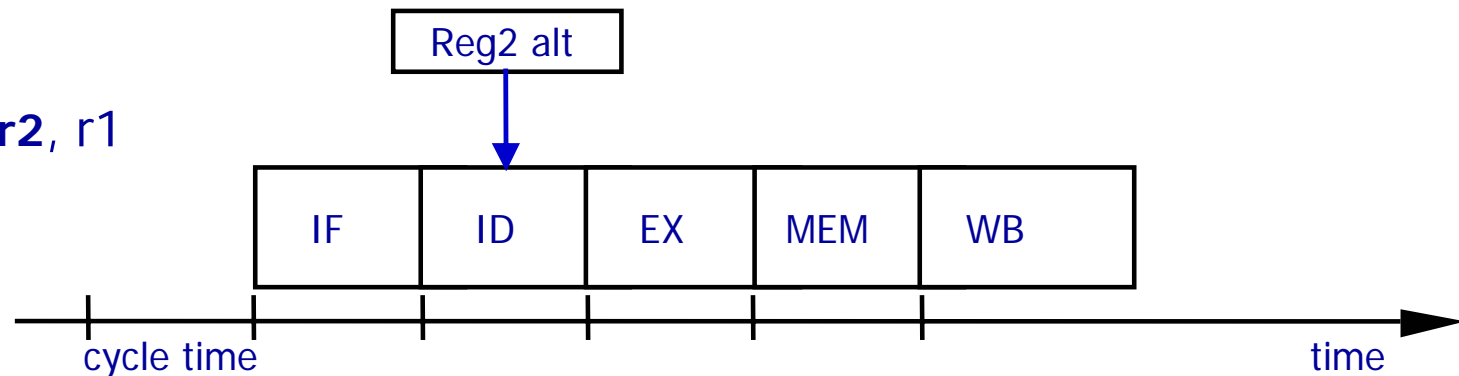
□ Software-Lösungen (*Compiler scheduling*)

- noop Befehle einfügen

add r2, r1, r2



mul r1, r2, r1



5.5.2 Lösungen für Datenkonflikte

□ Software-Lösungen (*Compiler scheduling*)

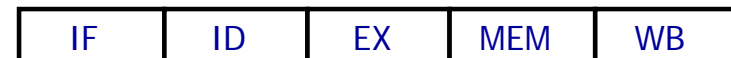
- **Befehlsanordnung** (*instruction scheduling* oder *pipeline scheduling*): Befehlsumordnungen, um noops zu entfernen

add r0, r1, r2

sub r3, r0, r4

mul r5, r6, r7

mul r8, r9, r10



5.5.2 Lösungen für Datenkonflikte

- **Hardware-Lösungen:** Konflikt muss per HW entdeckt werden!!
 - **Leerlauf der Pipeline:** Die einfachste Art, mit solchen Datenkonflikten umzugehen ist es, $Inst_2$ in der Pipeline für zwei Takte anzuhalten. Auch als **Pipeline-Sperrung** (*interlocking*) oder **Pipeline-Leerlauf** (*stalling*) bezeichnet.
 - **Forwarding:** Wenn ein Datenkonflikt erkannt wird, so sorgt eine Hardware-Schaltung dafür, dass der betreffende Operand nicht aus dem Universalregister, sondern direkt aus dem ALU-Ausgaberegister der vorherigen Operation in das ALU-Eingaberegister übertragen wird.
 - **Forwarding with interlocking:** Forwarding löst nicht alle möglichen Datenkonflikte auf.



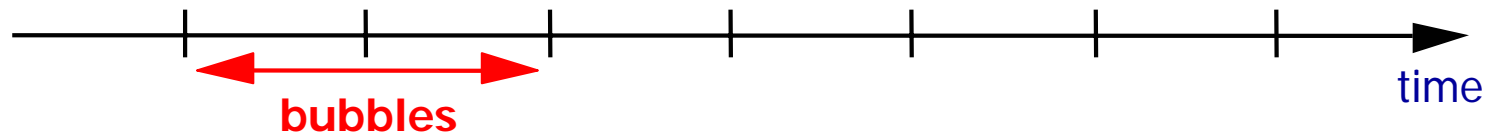
Hardware-Lösung durch Interlocking

add r2, r1, r2



Register Reg2

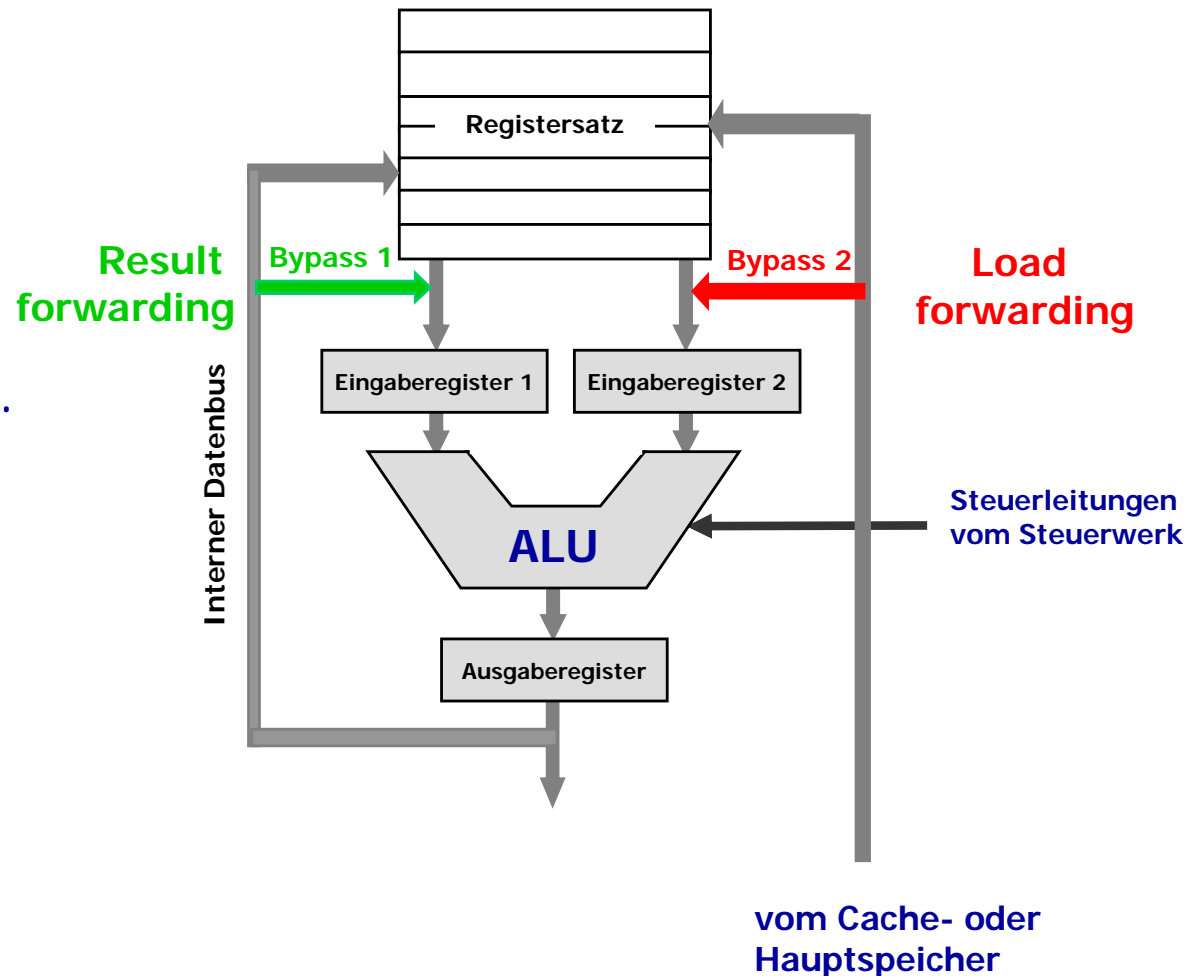
mul r1, r2, r1



Forwarding-Techniken

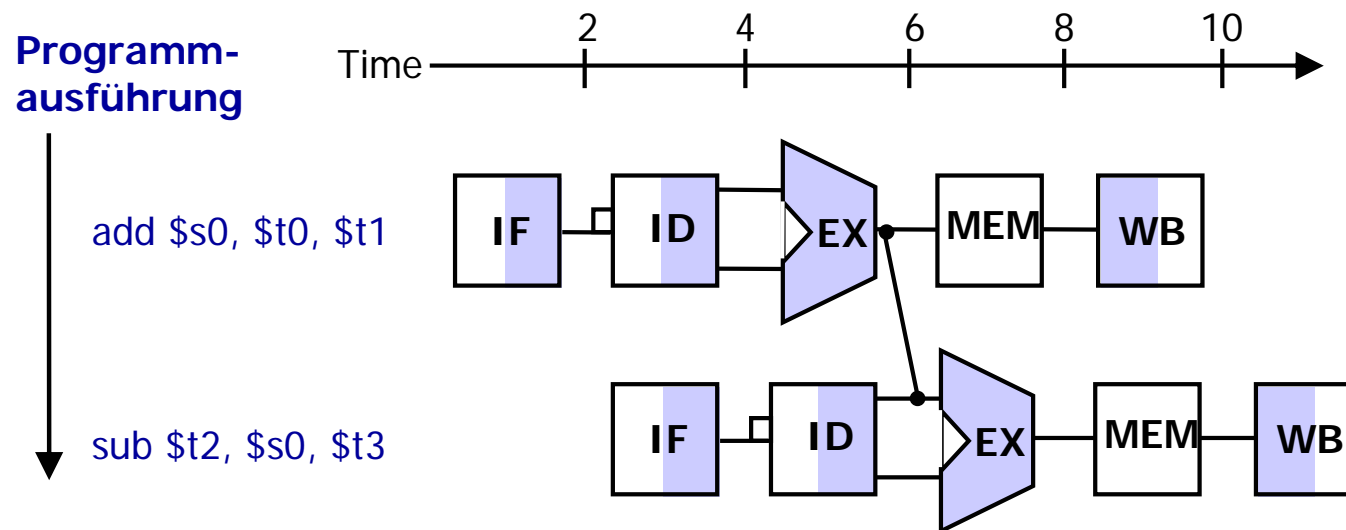
Forwarding:

- Rückführung des ALU-Ausgaberegister (*result forwarding*) bzw. des Ladewertregisters (*load forwarding*) auf die ALU-Eingaberegister
- Erhöhter Hardware- und Steuerungsaufwand



Forwarding-Techniken

Result Forwarding:



- Erhöhter Hardware- und Steuerungsaufwand für Forwarding-Logik und zusätzliche Datenpfade



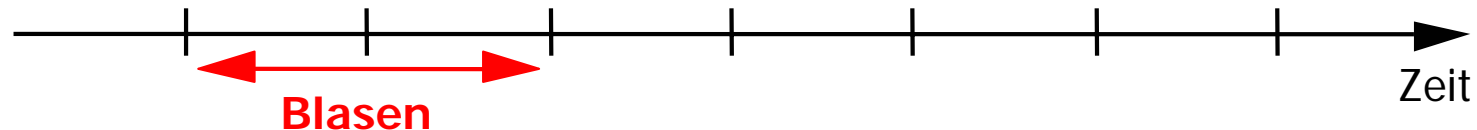
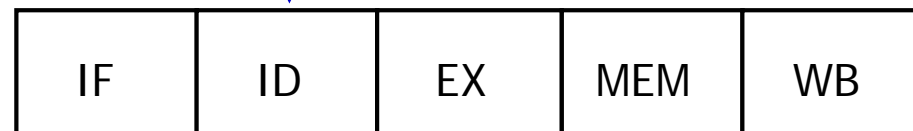
Leerlauf der Pipeline: Interlocking

`add r2, r1, r2`



Register r2

`mul r1, r2, r1`



Anhalten des `mul`-Befehls für zwei Takte

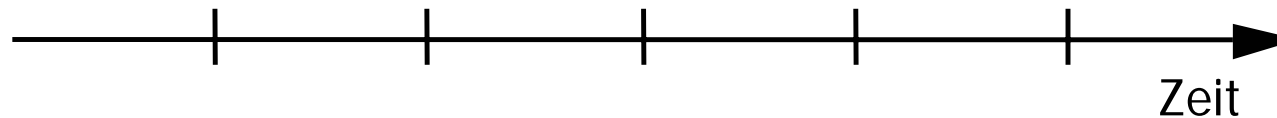


Hardware-Lösung durch *Forwarding*

`add r2, r1, r2`

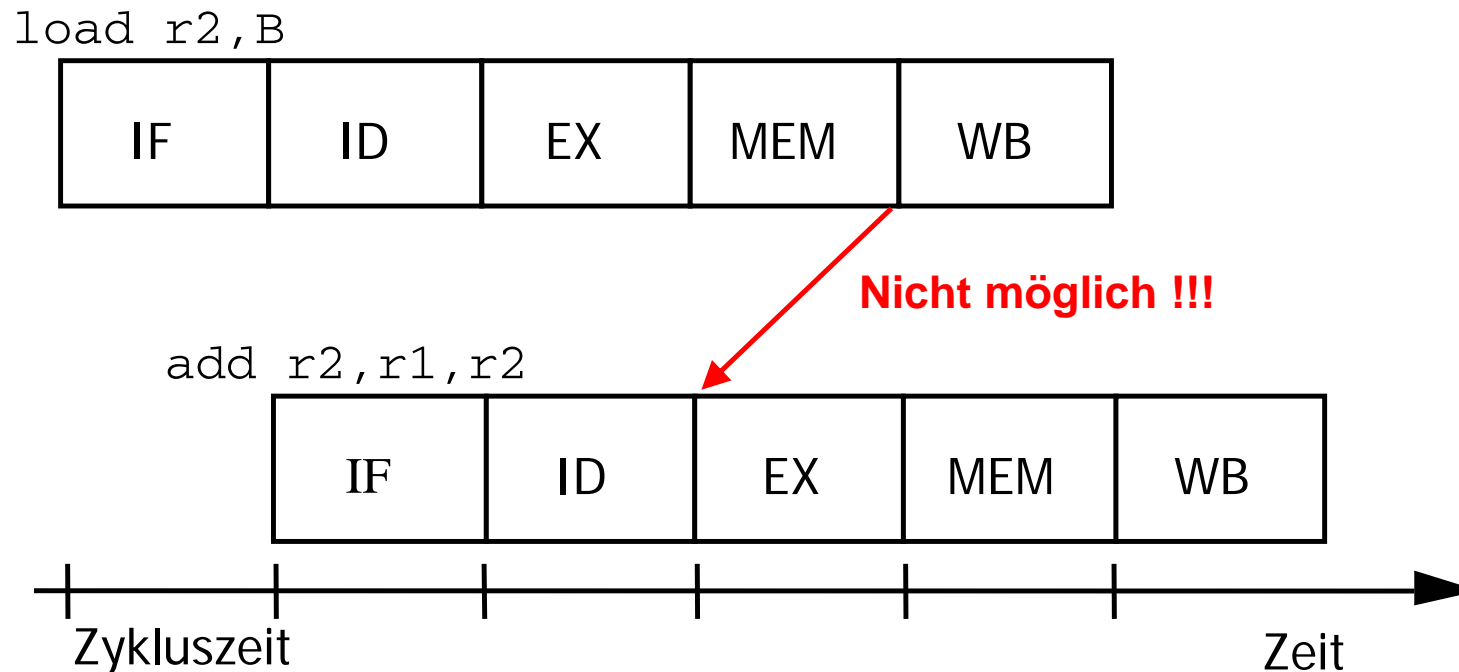


`mul r1, r2, r1`



Problem: Nicht alle Konflikte sind alleine durch Forwarding behebbar !!!

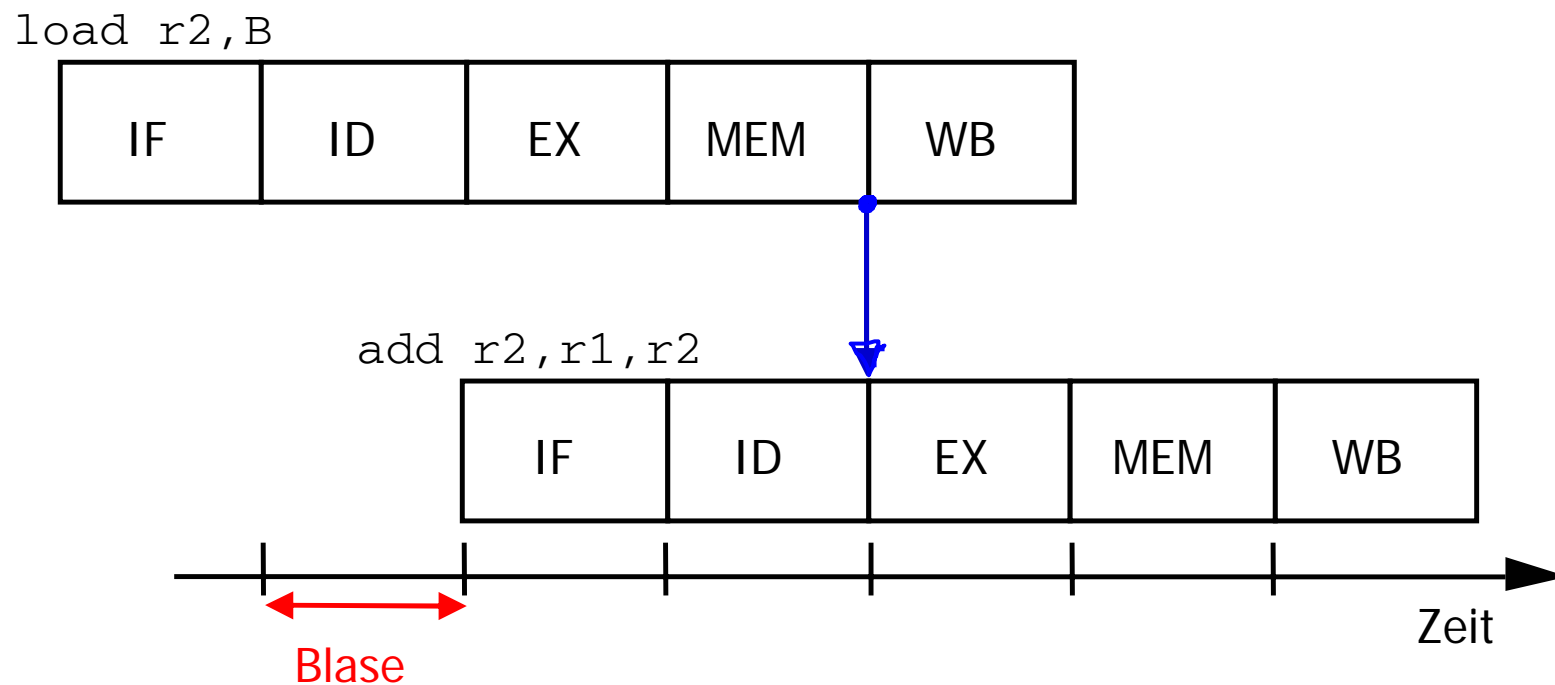
Beispiel: Speicherzugriff (z. B. Ladebefehl)



- EX-Stufe berechnet die effektive Adresse
- Der add-Befehl muss angehalten werden, bis die geladenen Daten im Ladewertregister der MEM-Stufe verfügbar sind



Lösung: Pipelineverzögerungen (bubble)



5.5.2 Lösungen für Datenkonflikte

□ Software-Lösung

➤ Aufgabe des Compilers:

- Erkennen von Datenkonflikten
- Einfügen von Leeroperationen (`noop`) nach jedem Befehl, der einen Konflikt verursacht oder verursachen kann.

➤ Statische Befehlsanordnung:

- Instruction Scheduling, Pipeline Scheduling
- Umordnen der Befehle des Programms (Code-Optimierung), um Leeroperationen zu eliminieren.



5.5.2 Lösungen für Datenkonflikte

- **Hardware-Lösungen:** Konflikt muss per HW entdeckt werden!!
 - **Leerlauf der Pipeline:** Die einfachste Art, mit solchen Datenkonflikten umzugehen ist es, $Inst_2$ in der Pipeline für zwei Takte anzuhalten. Auch als **Pipeline-Sperrung** (*interlocking*) oder **Pipeline-Leerlauf** (*stalling*) bezeichnet.
 - **Forwarding:** Wenn ein Datenkonflikt erkannt wird, so sorgt eine Hardware-Schaltung dafür, dass der betreffende Operand nicht aus dem Universalregister, sondern direkt aus dem ALU-Ausgaberegister der vorherigen Operation in das ALU-Eingaberegister übertragen wird.
 - **Forwarding with interlocking:** Forwarding löst nicht alle möglichen Datenkonflikte auf.



5.5.3 Ressourcenkonflikte

□ Ressourcenkonflikt:

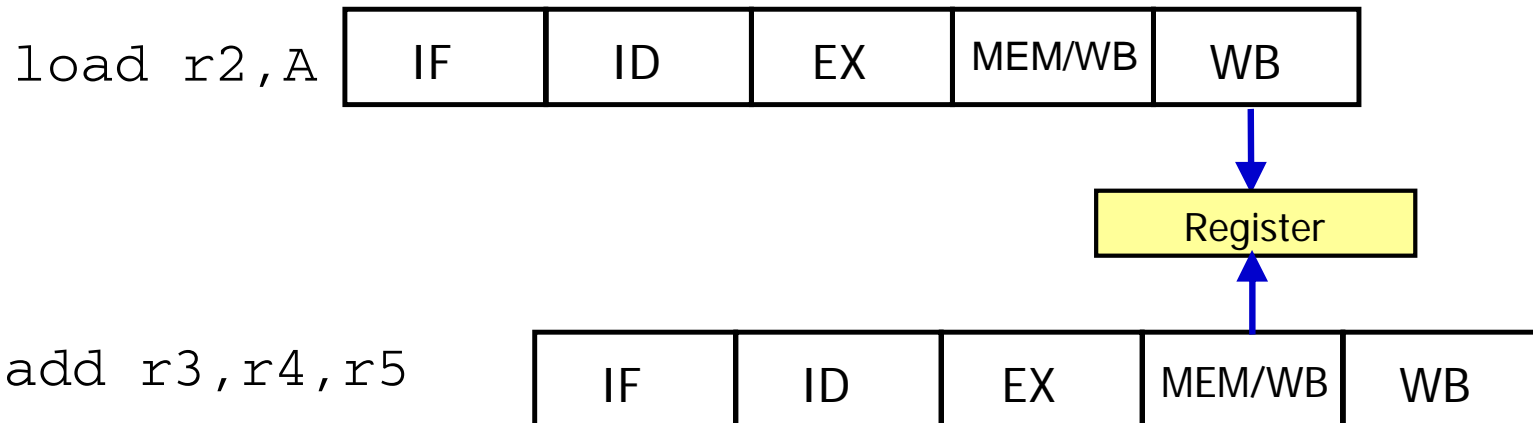
- Treten auf, wenn zwei oder mehr Befehle gleichzeitig dieselbe Ressource benötigen, auf die aber nur einmal zugegriffen werden kann
- Treten bei einer einfachen Pipeline, wie die DLX-Pipeline, nicht auf
- **Ziel beim Pipeline-Entwurf:** Ressourcenkonflikte möglichst zu vermeiden und dort, wo sie nicht vermeidbar sind, zu erkennen und behandeln



5.5.3 Ressourcenkonflikte

□ Ressourcenkonflikt:

- **Beispiel:** Prozessor mit veränderter DLX-Pipeline
 - Die MEM-Stufe ist in der Lage, bei einem Register-Register-Befehl die Ausgabe der ALU durch einen Schreibkanal auf den Registersatz zurückzuschreiben.
 - Im Beispiel greifen zwei Befehle gleichzeitig auf einen nur einfach vorhandenen Schreibeingang zu.



5.5.4 Lösungen von Ressourcenkonflikte

□ Lösungen (Hardware):

➤ Arbitrierung mit Interlocking

Arbitrierungslogik hält den Befehl, der den Konflikt verursacht → Verzögerung durch Leerzyklen

➤ Übertaktung:

Die Ressource, die den Konflikt hervorruft schneller als die übrigen Pipeline-Stufen takten

➤ Ressourcenreplizierung: Vervielfachung der Ressourcen

Beispiel: Registersatz mit mehreren Schreibkanälen



5.5.5 Steuerflusskonflikte

➤ **Programmsteuerbefehle:**

- die bedingten und die unbedingten Sprungbefehle,
- die Unterprogrammaufruf- und -rückkehrbefehle sowie
- die Unterbrechungsbefehle

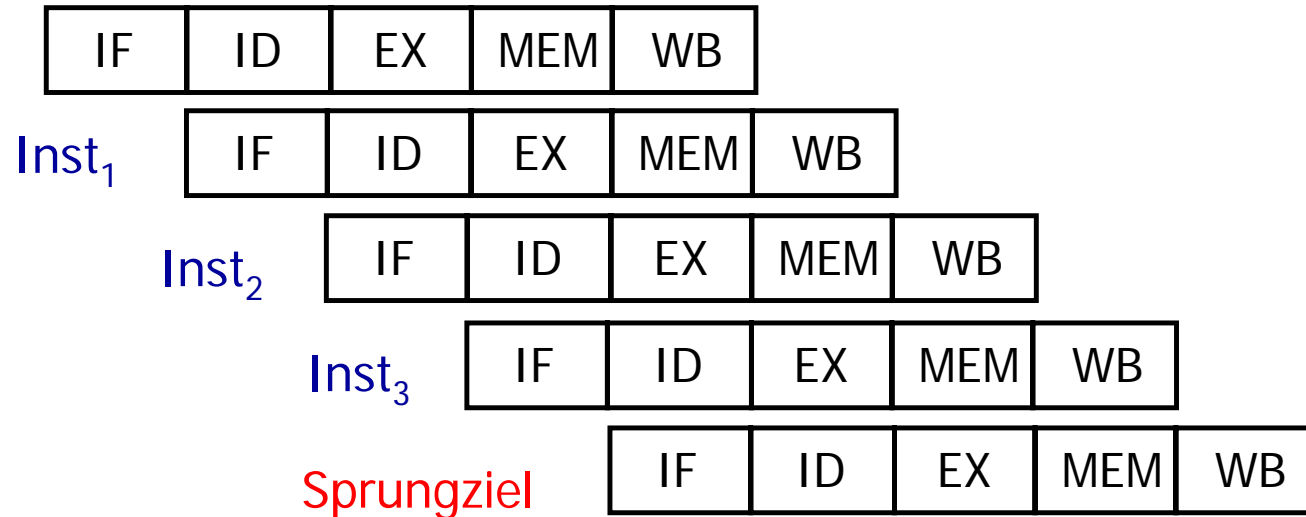
➤ **Steuerflussabhängigkeiten verursachen Steuerflusskonflikte in der DLX-Pipeline, da**

- der Programmsteuerbefehl erst in der ID-Stufe als solcher erkannt und damit bereits ein Befehl des falschen Programmpfades in die Pipeline geladen wurde.
- die Sprungzieladresse von der ALU erst in der EX-Stufe berechnet wird und
- der PC am Ende der MEM-Stufe ersetzt wird



Steuerflusskonflikte durch Verzweigung

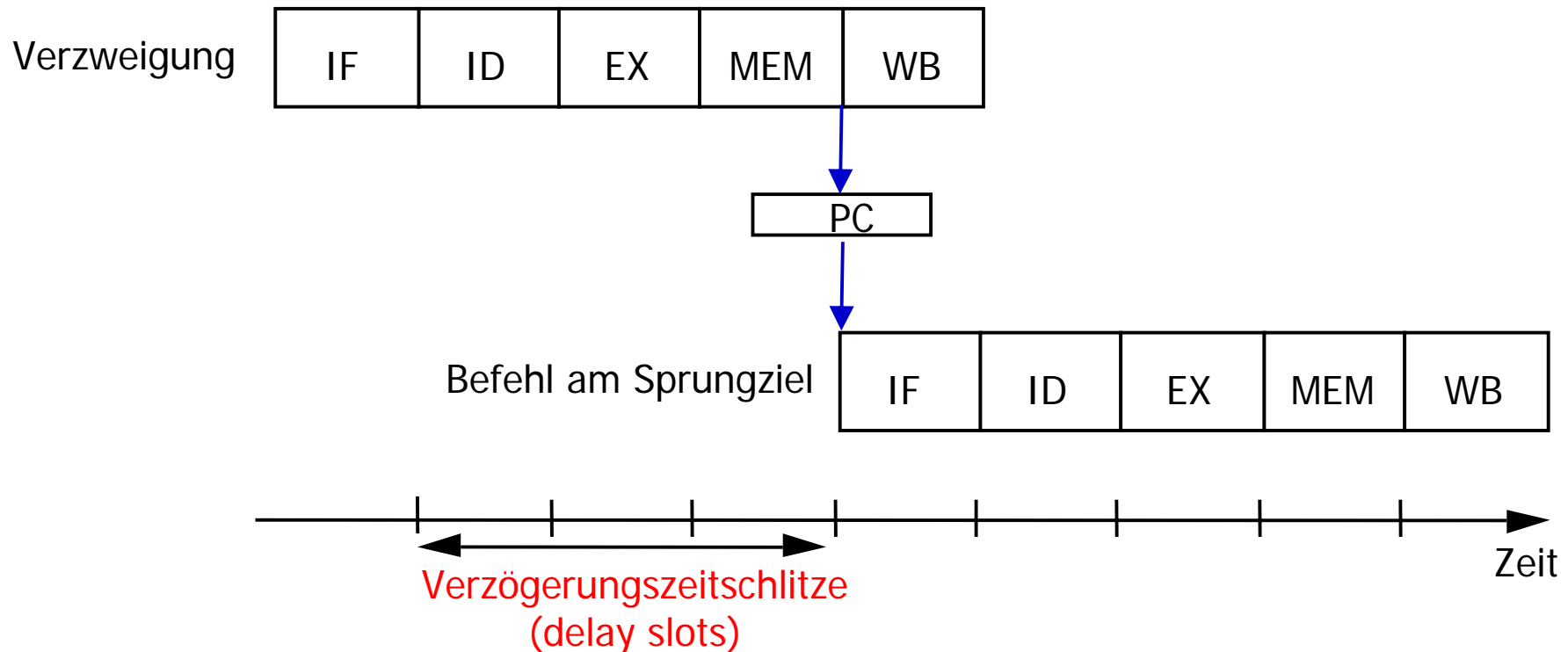
Sprungbefehl



- In der Pipeline befinden sich drei Befehle, die bereits geladen worden sind und wieder gelöscht werden müssen, wenn der Sprung ausgeführt wird.
- Bei einer Verzweigung kann erst nach **drei Taktzyklen** mit der Ausführung der korrekten Befehlsfolge gestartet werden.



Steuerflusskonflikte durch Verzweigung



- Nach einem genommenen bedingten Sprung ist eine Pipeline-Verzögerung durch Einfügen von Wartezyklen notwendig



Steuerflusskonflikte durch Verzweigung

Verringerung der Anzahl der Wartezyklen:

- Sprungrichtung so schnell wie möglich entscheiden und Zieladresse so schnell wie möglich berechnen.
- Berechnung in der ID-Stufe → nur noch ein Wartezyklus, aber dann → **Strukturkonflikt**
 - Die ALU kann nicht zur Berechnung der Sprungzieladresse benutzt werden, da sie vom vorhergehenden Befehl benötigt wird
- Dekodierung, Berechnung der Sprungzieladresse und Rückschreibung des PCs in einer Pipeline-Stufe → Kritischer Pfad in der Dekodierphase → Verlängerung der Zykluszeit!



Steuerflusskonflikte durch Verzweigung

- Trotz der vorgeschlagenen Pipeline-Reorganisation bleibt ein Wartezyklus (Verzögerungsphase).
- Bei der DLX-Pipeline werden Steuerflusskonflikte durch die Hardware **weder erkannt noch behandelt**, d. h. die drei Befehle hinter einem Sprungbefehl werden immer ausgeführt.
- **Techniken zur Konfliktauflösung**
 - Software-Techniken
 - Hardware-Techniken



Steuerflusskonflikte durch Verzweigung

□ Software-Lösung

- **Verzögerte Sprungtechnik (delayed branch technique)**
 - Ausfüllen der Verzögerungszeitslitze (delay slots) mit Leerbefehlen (`noop`)
 - DLX-Pipeline: Drei Leerbefehle nach jedem Programmsteuerbefehl



Steuerflusskonflikte durch Verzweigung

□ Software-Lösung

➤ Statische Befehlsanordnung:

- Der Compiler verschiebt Befehle, die in der logischen Programmreihenfolge vor dem Sprungbefehl liegen, in die Verzögerungszeitschlitze
- Nur dann möglich, wenn die verschobenen Befehle keinen Einfluss auf die Sprungrichtung haben
- Gibt es keine Befehle, die in die Verzögerungszeitschlitze verschoben werden können, müssen Leerbefehle eingefügt werden
- **Nachteil:** Code wird abhängig von der Pipeline-Struktur



Steuerflusskonflikte durch Verzweigung

□ Hardware-Lösung (Dynamische Techniken)

➤ Pipeline-Leerlauf:


- Einfachste und ineffizienteste Methode
- Hardware erkennt Verzweigungsbefehle (in der ID-Stufe) und lädt keine weiteren Befehle in die Pipeline, bis die Zieladresse berechnet und im Falle bedingter Sprungbefehle die Sprungentscheidung getroffen ist
- Der eine Befehl, der bereits ins Pipeline-Register der IF-Stufe geladen wurde, muss gelöscht werden.



Steuerflusskonflikte durch Verzweigung

□ Hardware-Lösung (Dynamische Techniken)

➤ Spekulation auf nicht genommene bedingte Sprünge:

- Einfachste Technik der statischen Sprungvorhersagen
 - Annahme: Sprung wird **nicht** „genommen“
 - Nachfolgende Befehle werden in die Pipeline geladen
 - Falls der Sprung „genommen“ wird, müssen die drei Befehle gelöscht werden (Pipeline flushing).
 - Falls der Sprung nicht „genommen“ wird, so können die drei Befehle ausgeführt werden.
 - **Nachteil:** Ineffizient, da bedingt durch die in Programmen häufig auftretenden Schleifen die Mehrzahl der Sprünge genommen wird
- 



Steuerflusskonflikte durch Verzweigung

□ Hardware-Lösung (Dynamische Techniken)

- Wie können „genommene“ bedingte Sprünge und alle anderen Programmsteuerbefehle, die immer zu Steuerflussänderung führen, behandelt werden?
 - Nur möglich, wenn in der IF-Stufe der richtige nachfolgende Befehl geladen werden kann.
 - Kleiner Pufferspeicher (Sprungzieladress-Cache, *branch target address cache*) in der IF-Stufe:
 - Speichert nach dem ersten Durchlauf der Befehlsfolge die Adresse des Programmsteuerbefehls und die Sprungzieladresse
 - Beim nächsten Auftreten desselben Programmsteuerbefehls kann die Pipeline sofort mit dem Befehl an der Zieladresse geladen werden.



Hyper-Pipeline beim Pentium 4

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
TC	Nxt IP	TC	Fetch	Drive	Alloc	Rename	Que	Sch	Sch	Sch	Disp	Disp	RF	RF	Ex	Flgs	Br Ck	Drive	

- 20 Stufen
- Bis zu 126 Instruktion werden gleichzeitig bearbeitet, davon 48 Load- und 24 Store-Operationen
- Kurze Stufen → Weniger Transistoren müssen geschaltet werden → Schneller



