

Kapitel 4

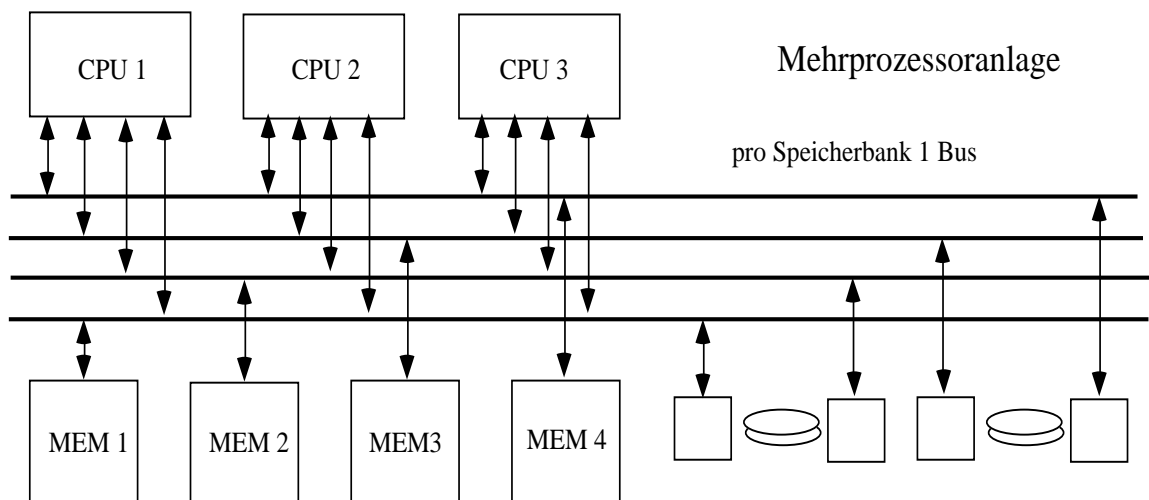
Mehrprozessoranlagen

4.1. Mehrere CPU's über Busse an Speicherbänke gekoppelt

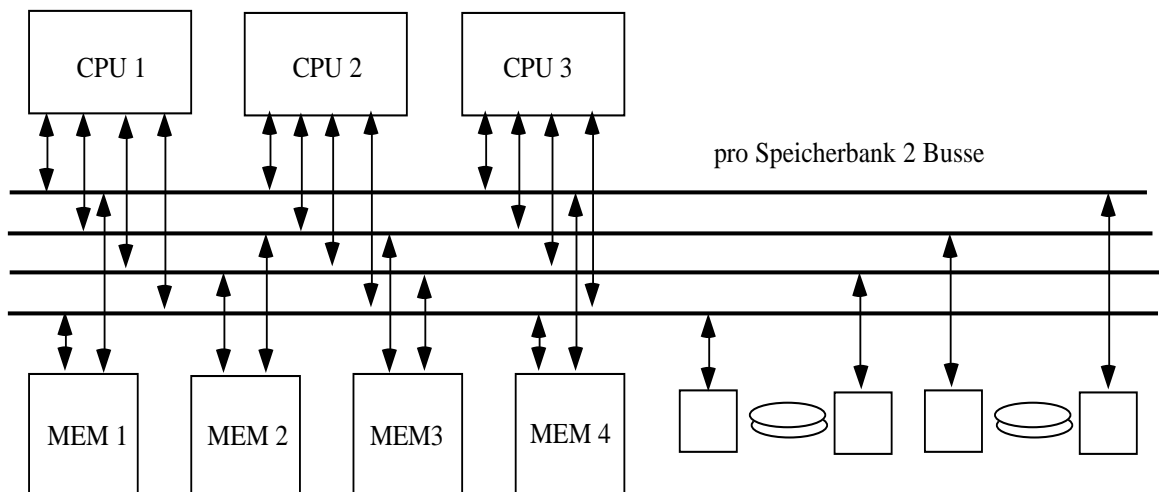
Die Erhöhung der Leistung eines Rechners kann durch Parallelschalten von CPU's und Speicherbänken erfolgen.

Eine solche Anlage hat eine relativ grobe Granularität: auf den CPU's laufen ganze Programme parallel. Das macht Sinn bei Time-Sharing-Anlagen, wo die Programme untereinander wenig Daten austauschen.

Den potentiellen Zugriffskonflikt auf die Speicherbänke entschärft man durch parallele Busse: jeder Speicherbank ist z. B. ein eigener Bus zugeordnet.



Gegebenenfalls hat eine Speicherbank auch Zugriff auf mehrere Busse, wobei ein Zugriffskonflikt durch ein "dual-ported memory" weitgehend ausgeschlossen werden kann.



Vorteile von Mehrprozessoranlagen:

- Massenspeicher können von mehreren CPU's gemeinsam genutzt werden.
- Bei unabhängigen Programmen beschleunigt sich die Rechnung etwa proportional der Prozessorzahl.

Nachteile von Mehrprozessoranlagen:

- Die Verteilung von Daten und Programmen ohne Gefahr gegenseitiger Blockaden ist bei mehreren Prozessoren problematisch.
- Bei Prozessoren mit Caches wird die Datenkonsistenz schwierig einzuhalten sein, wenn von mehreren Rechnern das gleiche Speichermodul angesprochen wird.

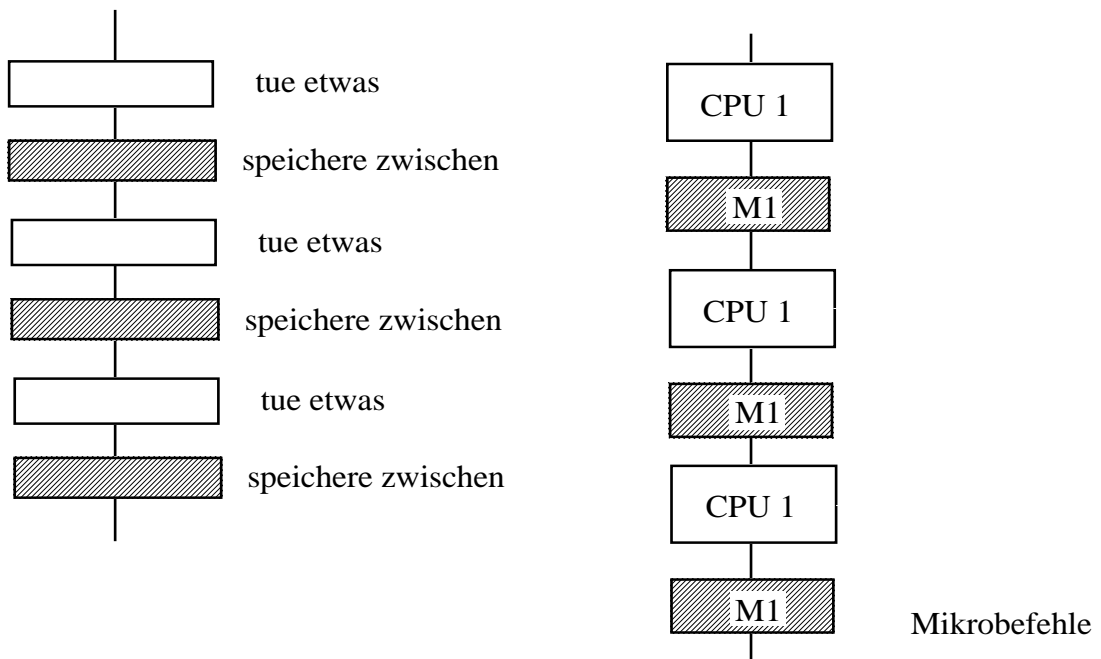
Doppelprozessoranlagen sind kommerziell erfolgreich.

Mehr als 4 Prozessoren werden nur in Einzelfällen eingesetzt.

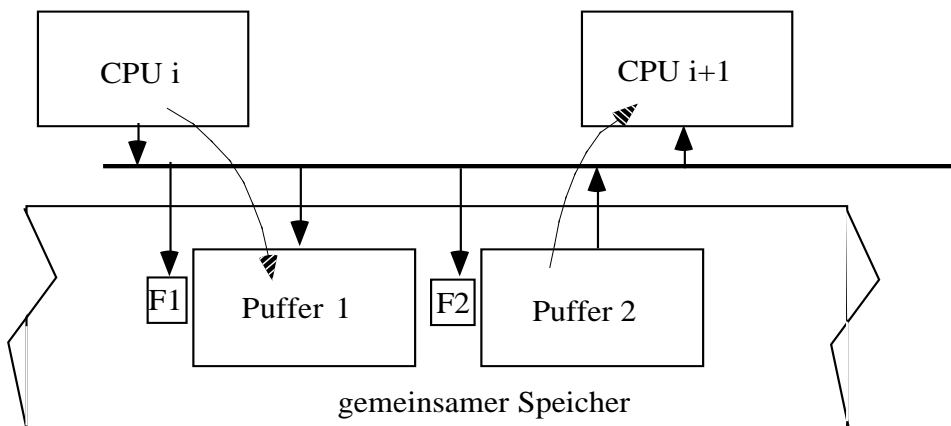
4.2. Mehrere Prozessoren in einer Pipeline

Sind umfangreiche Rechnungen sequentiell durchzuführen, kann der Einsatz von mehreren Prozessoren in einer Pipeline sinnvoll sein, die über einen gemeinsamen Speicher Daten austauschen.

Die Organisation der Rechnung muß dabei berücksichtigen, daß jeder Prozessor die gleiche Last trägt, um Wartezeiten bei der Synchronisation zu vermeiden.



Die Synchronisation erfolgt entweder zentral durch eine gemeinsame Taktung z. B. durch gemeinsame Mikrobefehle für alle Rechner oder durch das Loslassen einer gemeinsamen Meldeleitung: nur wenn alle Beteiligte das Ende einer Berechnungsphase gemeldet haben, liegt die Leitung auf H, und das ist das Signal zur Datenübernahme der Resultate der jeweils vorhergehenden Stufe. Um die Schreib- Lesevorgänge in den Speichern zu trennen wird eine Wechselpuffertechnik verwendet. Zu einem Paar von Processoren i und $(i+1)$ gehören zwei Pufferbereiche im Speicher und zwei Flags F1 und F2.



Im dargestellten Beispiel belegt CPU 1 F1 mit $(i, \text{ schreiben})$ und CPU 2 F2 mit $(i+1, \text{ lesen})$ und CPU 1 sichert sich damit exklusiven Schreibzugriff auf Puffer 1 bzw. CPU 2 Lesezugriff auf Puffer 2.

Damit ist der Zugriff auf Puffer 1 und 2 geregelt. Wenn alle CPU's sich synchronisiert haben wird getauscht: CPU i schreibt nach F2: $(i, \text{ schreiben})$ und CPU 2 nach F1: $(i+1, \text{ lesen})$. CPU 2 findet in Puffer 1 ihre Daten und CPU 1 kann ihre Ergebnisse in Puffer 2 einschreiben. Die dort vorhandenen Daten werden nicht mehr gebraucht.

Ein Beispiel einer Pipeline aus Rechnern ist der Graphik-Prozessor der SGI-Rechner aus Kapitel 2.

Die Rechner sind mikroprogrammiert und werden zentral getaktet. Da alle Rechner auf einem Chip sind, ist die zentrale Taktung sinnvoll: die Datenübernahme wird eh durch den Takt bestimmt.

4.3. Parallele Spezialprozessoren

Für spezielle Anwendungszwecke baut man Mehrprozessoranlagen mit Signalprozessoren z. B. für Bildverarbeitung, Signalkompression, digitale Filterung.

Auf einer Steckkarte sind 2 - 8 Prozessoren zu finden. Sie sind über Speicher gekoppelt.