UNIVERSITÄT DUISBURG ESSEN

MERCATOR School of Management

Kapazitätsplanung

(„Prozessdesign“)

Kapazitätsplanung bei Werkstattproduktion

Warteschlangensysteme

Werkstücke („Items") bzw. Kunden- oder Produktionsaufträge („Jobs") — bei Dienstleistungsproduktion sind das i. d. R. die Kunden/Abnehmer selbst — warten auf Bearbeitung bzw. auf Service.

Bei perfekter Planung („Scheduling") kann das Kapazitätsangebot exakt auf die -nachfrage abgestimmt werden. Dies gelingt aber auch nur unter besonders günstigen Bedingungen.

Hinzukommende stochastische Einflüsse führen unvermeidlich zu Warteschlangen.

- Die Kunden haben gewisse Servicewünsche, bzw. es wird eine bestimmte Erzeugnisvariante verlangt.
- Die Erfüllung der Servicewünsche verlangt eine gewisse Bedienzeit (Servicezeit, Bearbeitungszeit).
- Die Kundenaufträge kommen in gewissen Zeitabständen an, bzw. es herrscht asynchroner Materialfluss vor.

Zufällige Schwankungen ($\Longrightarrow$ Stochastizität) !

Aufbau von Warteschlangen

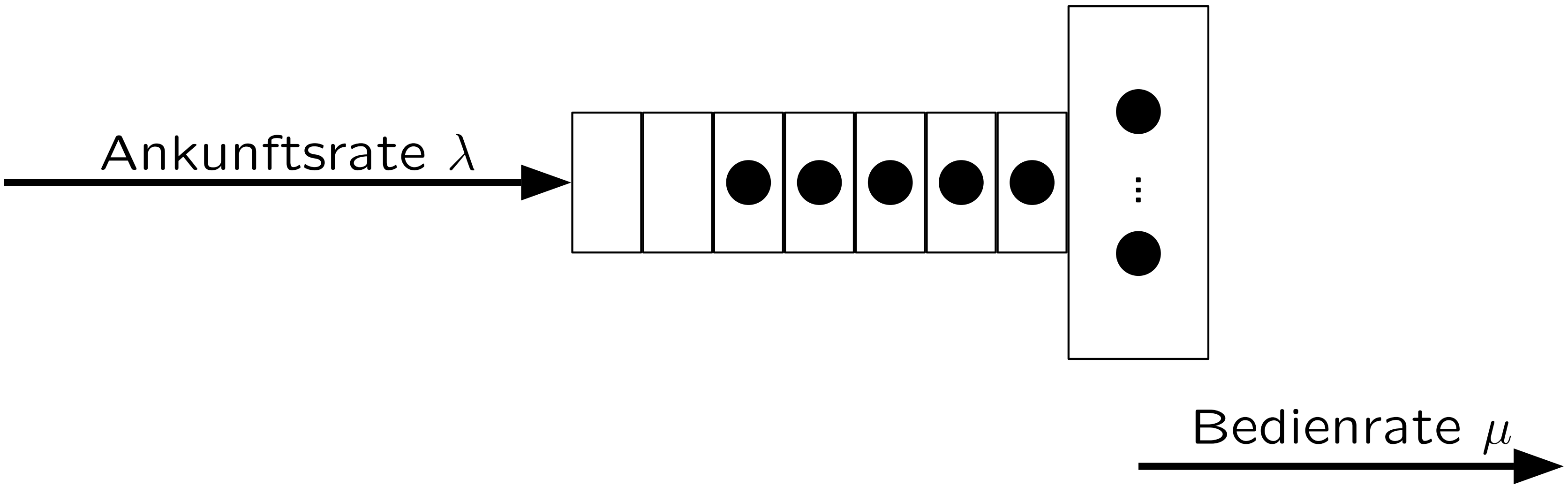
- Die Kunden haben ausgefallene, spezielle Servicewünsche.
- Die Erfüllung der Servicewünsche dauert außergewöhnlich lange.
- Es kommen gerade mehr Kunden an, als bedient werden können.

Abbau von Warteschlangen

- Die Kunden haben Standard-Servicewünsche.
- Die Erfüllung der Servicewünsche geht außergewöhnlich schnell.
- Es kommen gerade weniger Kunden an, als bedient werden können.

Das **Zusammenwirken dieser stochastischen Effekte** ist nicht exakt vorhersagbar!

Eine **zeitliche Koordination** von Servicewünschen und entsprechendem Kapazitätsangebot der Serviceeinrichtung ist nicht möglich!


Ankunftsrate λ
Bedienrate μ

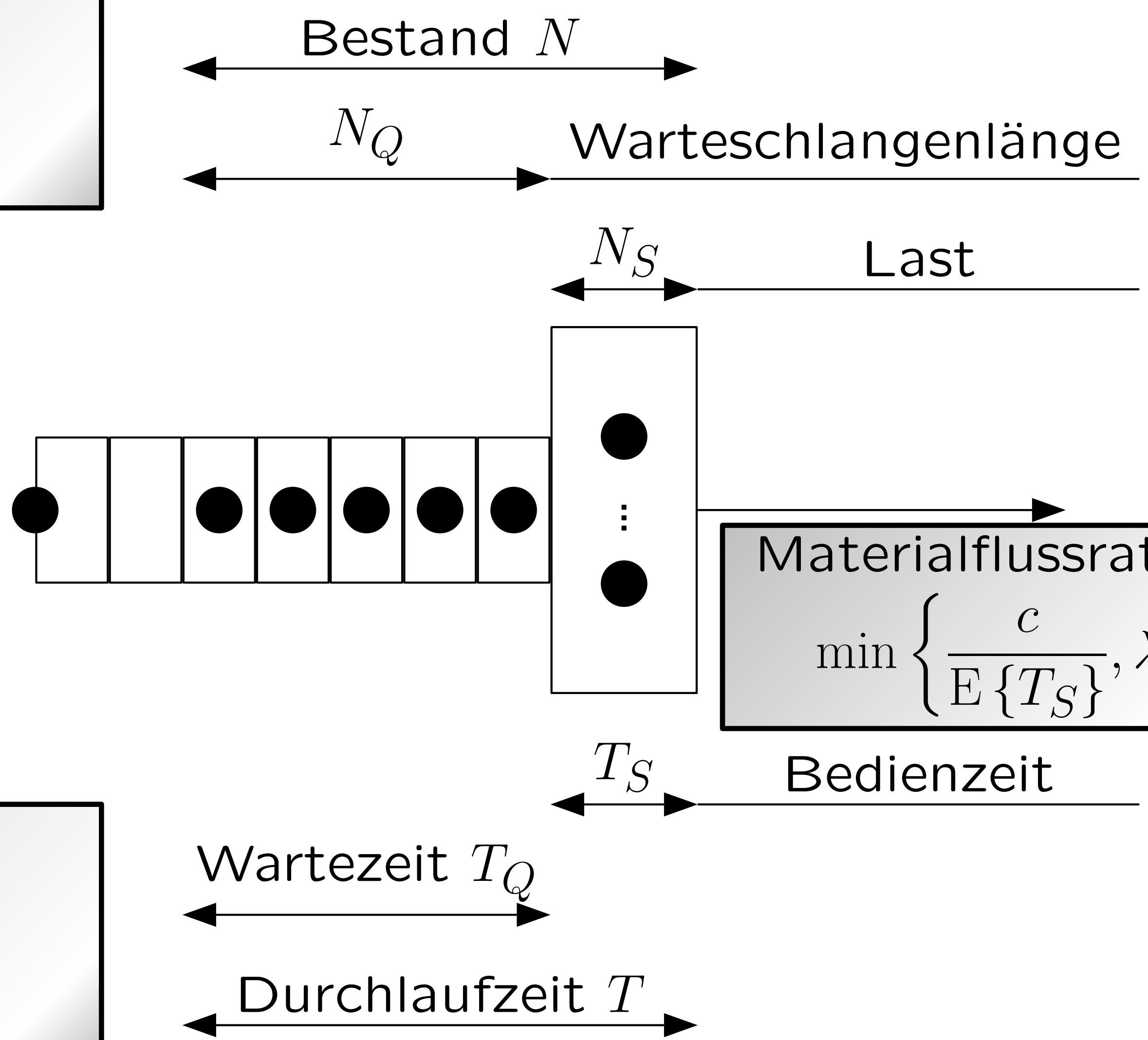
Anzahl Kunden im System (= Bestand)

$$L := \mathrm{E}\{N\}$$

Zwischenankunftszeit T_A

Materialflussrate

$$\lambda := \min\left\{\frac{1}{\mathrm{E}\{T_A\}}, \tilde{\lambda}\right\}$$



Materialflussrate

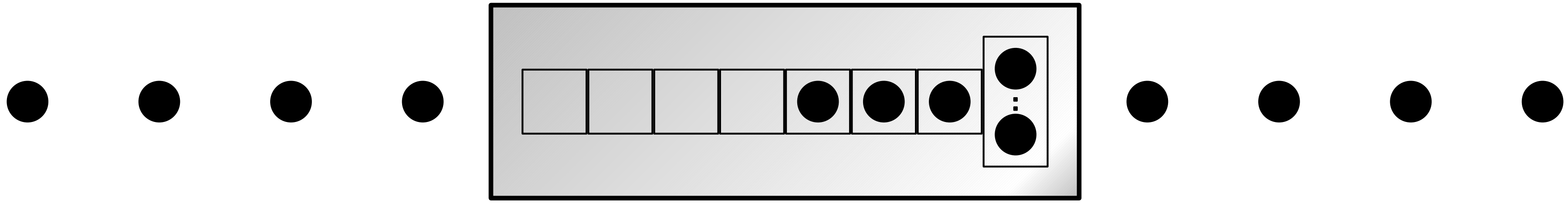
$$\min\left\{\frac{c}{\mathrm{E}\{T_S\}}, \lambda\right\}$$

Zeit im System (= Durchlaufzeit)

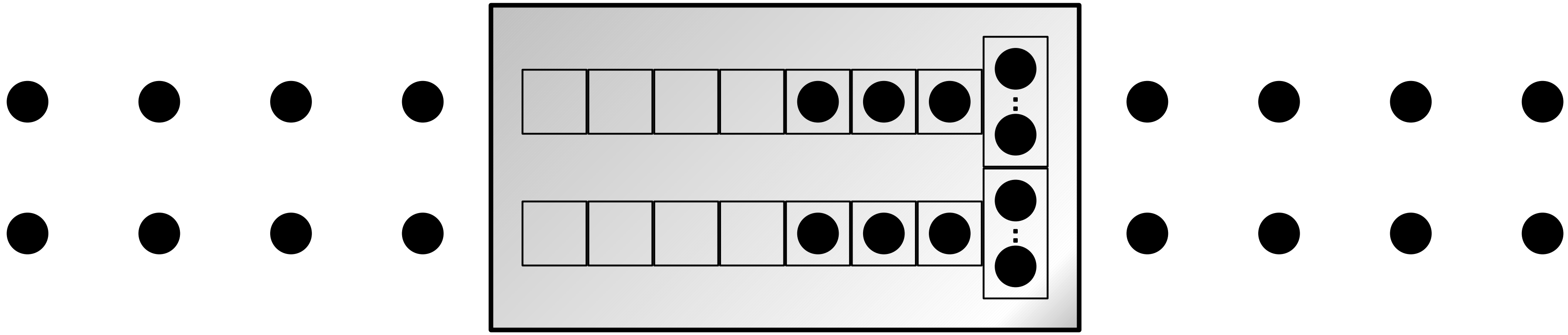
$$W := \mathrm{E}\{T\}$$

Littles Gesetz

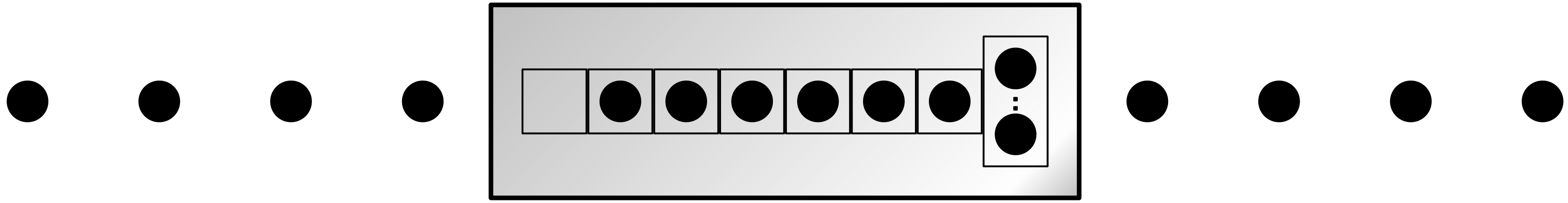
Bestand L = Materialflussgeschwindigkeit $\lambda \cdot$ Durchlaufzeit W

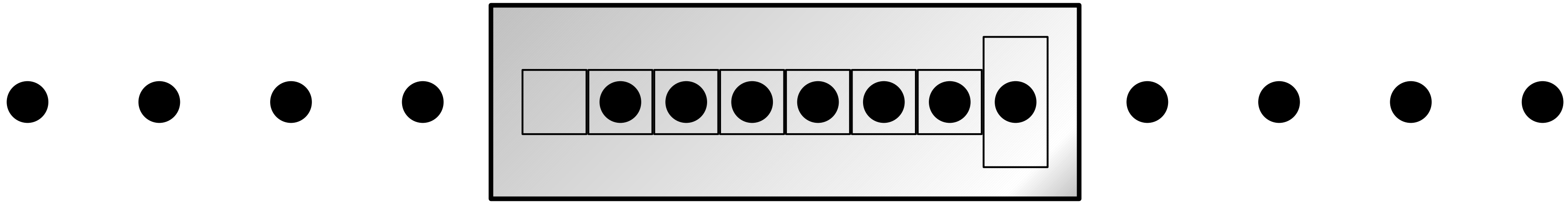


Doppelter Durchsatz bei gleicher Durchlaufzeit ($\rightarrow$ doppelter Bestand):



Gleicher Durchsatz bei erhöhter Durchlaufzeit ($\rightarrow$ gestiegener Bestand):





Auslastung (Workload Rate) (bei nur einem Server)

$$\rho = \mathrm{E}\left\{T_S\right\} \cdot \frac{1}{\mathrm{E}\left\{T_A\right\}} = \frac{\lambda}{\mu}$$

Bedienrate (Service Rate)

$$\mu = \frac{1}{\mathrm{E}\left\{T_S\right\}}$$

Ankunftsrate (Arrival Rate)

$$\lambda = \frac{1}{\mathrm{E}\left\{T_A\right\}}$$

[Spezialfall:]

- exponentialverteilte Zwischenankunftszeiten (Poisson-Ankunftsprozess)
  - Anzahl X ankommender Werkstücke ist poissonverteilt mit Rate λ

$$\mathrm{P}[X=1] = \mathrm{P}[N(t+\Delta t) - N(t) = 1] \approx \lambda\,\Delta t$$

- exponentialverteilte Bearbeitungszeiten
  - Anzahl Y fertiger Werkstücke (wenn n Server arbeiten) ist poissonverteilt mit Rate $n \cdot \mu$
- 1 Server

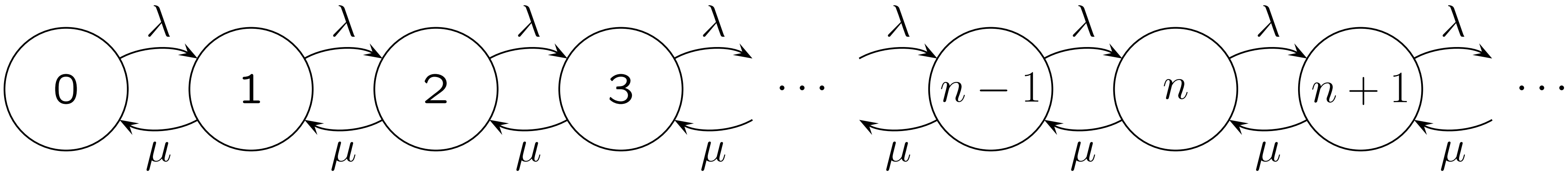
$$\mathrm{P}[Y=1] = \mathrm{P}[N(t+\Delta t) - N(t) = -1] \approx \mu\,\Delta t$$

$\Longrightarrow$ M/M/1-Warteschlangensystem

- unbeschränkter Warteraum, unbeschränkte Kundenquelle, FCFS-Warteschlangendisziplin (FCFS = first come first served)

M/M/1-Warteschlangensystem

Bestandsentwicklung



Zustandswahrscheinlichkeiten

$$
\begin{aligned}
P_n(t+\Delta t) = {} & P_n(t) \cdot \bigl(1-\lambda\,\Delta t\bigr) \cdot \bigl(1-\mu\,\Delta t\bigr) \\
& + P_{n+1}(t) \cdot \bigl(1-\lambda\,\Delta t\bigr) \cdot \bigl(\mu\,\Delta t\bigr) \\
& + P_{n-1}(t) \cdot \bigl(\lambda\,\Delta t\bigr) \cdot \bigl(1-\mu\,\Delta t\bigr) \\
& + P_n(t) \cdot \bigl(\lambda\,\Delta t\bigr) \cdot \bigl(\mu\,\Delta t\bigr)
\end{aligned}
$$

Auslastung

$$\rho = \frac{\lambda}{\mu} = \lambda \cdot \mathrm{E}\{T_S\}$$

Zustandswahrscheinlichkeiten

$$P_n = \mathrm{P}[N = n] = P_{n-1} \cdot \rho = P_0 \cdot \rho^n \qquad (n = 1, 2, 3, \ldots)$$

$$P_0 = \mathrm{P}[N = 0] = 1 - \rho$$

erwarteter Bestand im System

$$L = \mathrm{E}\{N\} = \frac{\rho}{1 - \rho}$$

Littles Gesetz: $L = \lambda \cdot W$

erwartete Durchlaufzeit

$$W = \mathrm{E}\{T\} = \frac{L}{\lambda} = \frac{1}{\mu \cdot (1 - \rho)}$$

Auslastung

$$\rho = \frac{\lambda}{\mu} = \lambda \cdot \mathrm{E}\left\{T_S\right\}$$

Zustandswahrscheinlichkeiten

$$P_n = \mathrm{P}\left[N = n\right] = P_{n-1} \cdot \rho = P_0 \cdot \rho^n \qquad (n = 1, 2, 3, \ldots)$$

$$P_0 = \mathrm{P}\left[N = 0\right] = 1 - \rho$$

erwartete Warteschlangenlänge

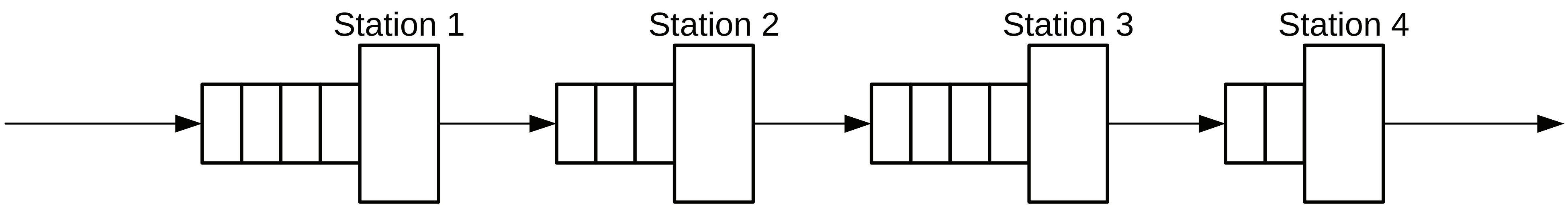
$$L_Q = \mathrm{E}\left\{N_Q\right\} = L - \mathrm{E}\left\{N_S\right\} = \frac{\rho^2}{1 - \rho}$$

Littles Gesetz: $L_Q = \lambda \cdot W_Q$

erwartete Wartezeit

$$W_Q = \mathrm{E}\left\{T_Q\right\} = \frac{L_Q}{\lambda} = \frac{\rho}{\mu \cdot (1 - \rho)}$$

einheitlicher Materialfluss



Kapazitätsplanung bei Fließproduktion

Beispiel Arbeitselemente

Arbeitselement Nr. i	1	2	3	4	5	6	7	8
Elementzeit t_i [Zeiteinheiten]	3	1	2	5	4	4	7	1

Vorranggraph (Reihenfolgerestriktionen)

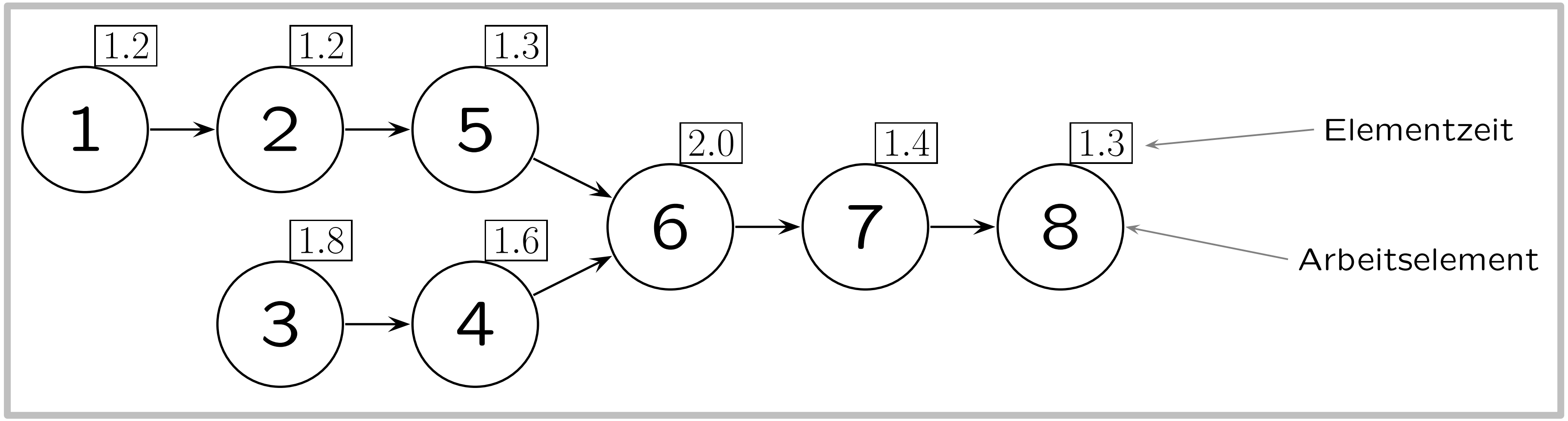
$$1 \rightarrow 4 \rightarrow 7 \rightarrow 5 \rightarrow 2 \rightarrow 3 \rightarrow 6 \rightarrow 8$$

(Beispiel aus Günther/Tempelmeier (2016))

Beispiel Arbeitselemente

Arbeitselement Nr. i	1	2	3	4	5	6	7	8
Elementzeit t_i [Zeiteinheiten]	1.2	1.2	1.8	1.6	1.3	2.0	1.4	1.3

Vorranggraph (Reihenfolgerestriktionen)



(Beispiel aus Günther/Tempelmeier (2016))

Planungsprobleme:

- **Leistungsabstimmung**
  - Arbeitsanalyse hinsichtlich
    - Arbeitselemente (elementare Arbeitsgänge)
    - Elementzeit (Arbeitsgangdauer)
    - Produktmix (Varianten des Grundprodukts)
    - Vorranggraph (technologisch bedingte Reihenfolgerestriktionen)
  - Entscheidungsproblem
    - Minimierung der Anzahl benötigter Bearbeitungsstationen
    - Aufbau der einzelnen Bearbeitungsstationen: Zuordnung der Arbeitselemente zu den Stationen („Arbeitsinhalt“)
    - Restriktionen:
      - vorgegebene Produktionsrate bzw. Taktzeit
      - Unteilbarkeit der Arbeitselemente
      - z. T. technologisch zwingende Bearbeitungsreihenfolge

Leistungabstimmung

[Zunächst unterstellenw ir deterministische Bedingungen:]

- keine Maschinenausfälle
- keine schwankenden Bearbeitungszeiten
- keine variantenabhängigen Elementzeiten
- synchroner Materialfluss

= deterministische Elementzeiten

= 100 %ig zuverlässige Transferstraßen

$\Longrightarrow$ „Klassische Fließbandabstimmung“

Das klassische Fließbandabstimmungsproblem (Simple Assembly-Line Balancing Problem)

Ein Optimierungsmodell zur klassischen Fließbandabstimmung

MERCATOR School of Management

Modell SALBP

Was muss festgelegt werden — Entscheidungsvariablen:

$y_m \in \{0;1\}$ ... Wieviel Stationen werden benötigt?

$$y_m = \begin{cases} 1 & \text{, wenn die } m\text{te Station benötigt wird} \\ 0 & \text{sonst} \end{cases}$$

$x_{im} \in \{0;1\}$ ... Welche Arbeitselemente werden an den einzelnen Stationen ausgeführt?

$$x_{im} = \begin{cases} 1 & \text{, wenn Arbeitselement } i \text{ der Station } m \text{ zugeordnet wird} \\ 0 & \text{sonst} \end{cases}$$

Modell SALBP

Was ist gegeben — Indexmengen:

I ... Anzahl Arbeitselemente

M ... maximale Anzahl Stationen

$\mathcal{N}_i$ ... Menge der direkten Nachfolger des Arbeitselements i

Was ist gegeben — Daten:

C ... Taktzeit

t_i ... Elementzeit (Arbeitsgangdauer) des Arbeitselements i

Modell SALBP

Minimiere die Anzahl benötigter Stationen: $Z = \sum_{m=1}^{M} y_m$

unter Berücksichtigung von

Zuordnungsrestriktionen
Vollständige und eindeutige Zuordnung der Arbeitselemente $i = 1, 2, \ldots, I$:

$$\sum_{m=1}^{M} x_{im} = 1$$

Taktzeitrestriktionen: Obergrenze an allen Stationen $m = 1, 2, \ldots, M$:

$$\sum_{i=1}^{I} t_i \cdot x_{im} \leq C \cdot y_m$$

Reihenfolgerestriktionen in bezug die Arbeitselemente $i = 1, 2, \ldots, I - 1$:

$$\sum_{m=1}^{M} m \cdot x_{im} \leq \sum_{m=1}^{M} m \cdot x_{jm} \quad \text{(jeweils für die Nachfolgearbeitsgänge } j \in \mathcal{N}_i\text{)}$$

Äquivalente Zielsetzungen:

- Minimiere die **Anzahl Stationen**:

$$Z$$

- Minimiere die **Durchlaufzeit** eines Werkstücks:

$$Z \cdot C$$

- Minimiere die Summe der **Leerzeiten**:

$$\sum_{m=1}^{Z} \left(C - \sum_{i=1}^{I} t_i \cdot x_{im} \right) = Z \cdot C - \sum_{i=1}^{I} t_i$$

- Maximiere die **Auslastung** (den Bandwirkungsgrad):

$$1 - \frac{Z \cdot C - \sum_{i=1}^{I} t_i}{Z \cdot C} = \frac{\sum_{i=1}^{I} t_i}{Z \cdot C}$$

Heuristische Lösungsverfahren

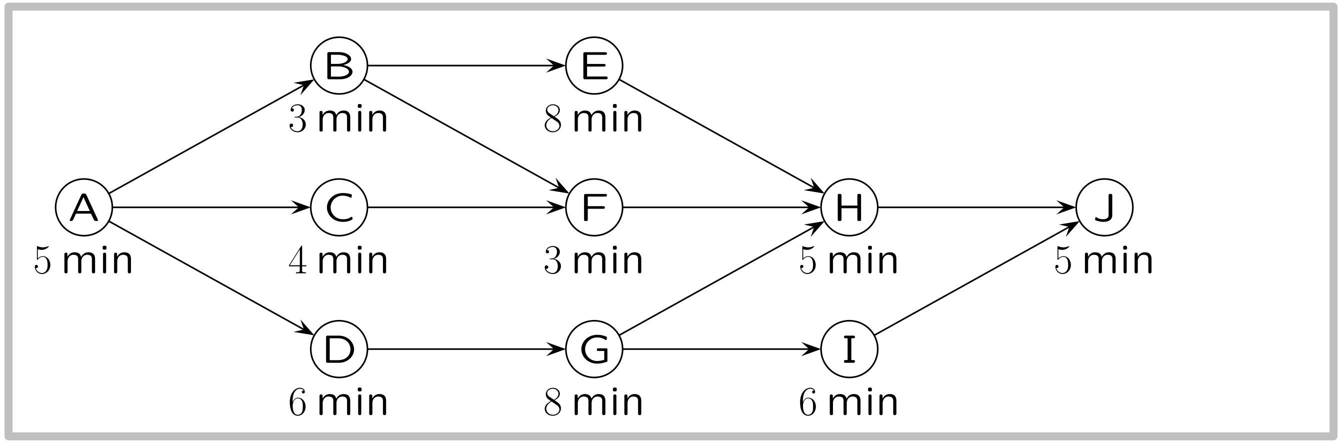
Bilde sukzessive die Stationen durch Zuordnung möglichst vieler Arbeitselemente unter Beachtung der

- Reihenfolgerestriktionen,
- Taktzeitrestriktion !

Wähle aus mehreren möglichen Arbeitselementen gemäß einem **Prioritätsregelverfahren** z. B. dasjenige mit

- der größten Anzahl Nachfolger,
- der längsten Elementzeit,
- der kürzesten Elementzeit,
- dem höchsten Positionswert !
  (= Summe der Elementzeit mit den Elementzeiten der Nachfolger)

Beispiel **Klassische Fließbandabstimmung**



vorgegebene Produktionsrate

$X = 42$ Stück pro Schicht, d. h. 42 Stück in $T = 462$ [Minuten]

maximale (größtzulässige) Taktzeit

(= Kehrwert der vorgegebenen Produktionsrate)

$$C = \frac{T}{X} = \frac{462\,[\text{Minuten}]}{42\,[\text{Stück}]} = 11 \text{ Minuten pro Stück}$$

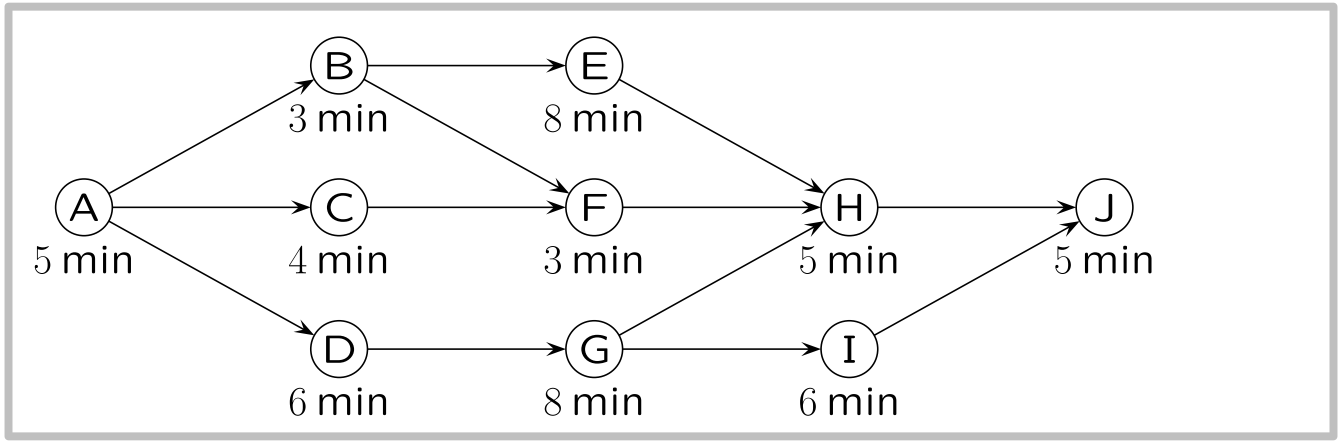
Mindestanzahl (theoretisch minimale Anzahl) an Stationen

$$M_{\min} = \frac{\sum \text{Elementzeiten}}{C} = \frac{53}{11} = 4.8\bar{1} \implies \text{aufgerundet: } M_{\min} = \lceil 4.8\bar{1} \rceil = 5$$

Beispiel **10 Arbeitselemente, Taktzeit 11 Minuten**

Station	Arbeitselemente einplanbar nach Reihenfolge	Arbeitselemente einplanbar nach Taktzeit	Arbeitselemente ausgewählt	Elementzeit [min]	Stationszeit [min]	Restzeit [min]
I	A	A	A	5	5	6
	B, C, D	B, C, D	B mehr Nachfolger als C, ansonsten willkürlich	3	8	3
II	C, D, E	C, D, E	D hat die meisten Nachfolger	6	6	5
	C, E, G	C nur C passt noch	C	4	10	1
III	E, F, G	E, F, G	G hat die meisten Nachfolger	8	8	3
	E, F, I	F nur F passt noch	F	3	11	0
IV	E, I	E, I	E mehr Nachfolger als I	8	8	3
V	H, I	H, I	H willkürlich	5	5	6
	I	I	I	6	11	0
VI	J	J	J	5	5	6

Beispiel **Klassische Fließbandabstimmung**



Bandwirkungsgrad

$$U = 1 - \frac{\sum \text{Leerzeiten}}{\text{Gesamtdurchlaufzeit}} = 1 - \frac{3+1+0+3+0+6}{6 \cdot 11} = 80.\overline{30}\,\%$$

Stochastisch schwankende Bearbeitungszeiten führen zu zufälligen Wartezeiten vor den Stationen.

Puffer sind Warteräume/Lager für Werkstücke zwischen zwei Stationen.

Leistungsmindernde Effekte bei beschränkten Puffern:

- Starving: Eine Station kann nicht mit der nächsten Bearbeitung beginnen, weil der davorliegende Puffer leer ist.
- Blocking: Eine Station kann nicht mit der nächsten Bearbeitung beginnen, weil der nachfolgende Puffer voll ist und fertige Werkstücke nicht weitertransportiert werden können.

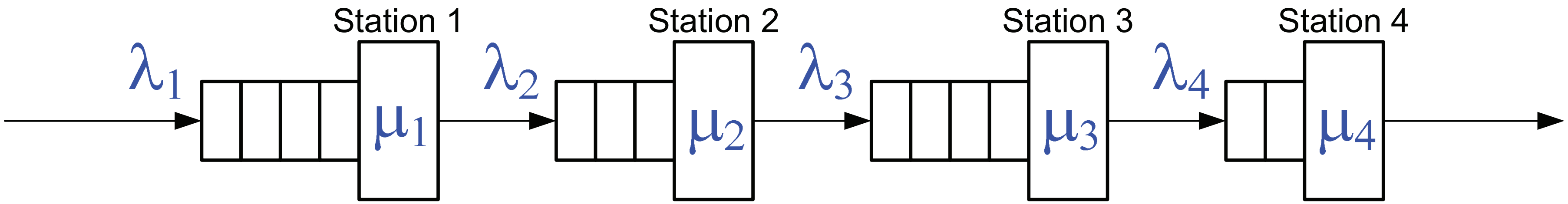
Effizienzüberlegungen:

- Das Ausmaß von Blocking und Starving verkleinert sich mit der Puffergröße. ($\Rightarrow$ höhere Produktionsrate)
- Pufferbestände verursachen Lagerkosten.
- Technische Lösungen für Puffereinrichtungen erfordern Investitionsaufwand.

Planungsprobleme:

- **Leistungsabstimmung**
  - Minimierung der Anzahl benötigter Bearbeitungsstationen
  - Aufbau der einzelnen Bearbeitungsstationen: Zuordnung der Arbeitselemente zu den Stationen („Arbeitsinhalt“)
  - Restriktionen:
    - vorgegebene Produktionsrate bzw. Taktzeit
    - Unteilbarkeit der Arbeitselemente
    - z. T. technologisch zwingende Bearbeitungsreihenfolge
- **Pufferallokation**
  - Verteilung einer minimalen Anzahl Pufferplätze zwischen die Stationen, so dass die angestrebte Systemleistung (Produktionsrate) erreicht wird
- **Leistungsanalyse**

Leistungsanalyse bei stochastischen Bearbeitungszeiten



Betrachtung der Stationen als eine Reihe von Warteschlangensystemen:

- μ_m ... Bearbeitungsrate der Station m,
  b_m ... mittlere Bearbeitungszeit eines Werkstücks an der Station m
  $\Longrightarrow \mu_m = \frac{1}{b_m}$
- λ_m ... Ankunftsrate von Werkstücken an der Station m
- unbeschränkte Puffer,
  d. h. Outputrate an Station m = $\min\{\lambda_m, \mu_m\} = \lambda_{m+1}$
- $\lambda_1 = \lambda$ ... Ankunftsrate von Werkstücken aus dem Lager für Vorprodukte
- Produktionsrate

$$X = \min\{\lambda, \mu_1, \mu_2, \ldots, \mu_M\}$$

[Spezialfall:]

- exponentialverteilte Zwischenankunftszeiten (Poisson-Ankunftsprozess)
- exponentialverteilte Bearbeitungszeiten, d. h. Stationszeiten

Auslastung der Station m: $(m = 1, 2, \ldots, M)$

$$\rho_m = \frac{\lambda_m}{\mu_m} = \lambda_m \cdot b_m$$

Zustandswahrscheinlichkeiten für die Station m: $(m = 1, 2, \ldots, M)$

$$P_{n_m} = (1 - \rho_m) \cdot \rho_m^{n_m} \qquad (n_m = 0, 1, 2, \ldots)$$

mittlerer Bestand *an* der Station m: $(m = 1, 2, \ldots, M)$

$$L_m = \frac{\rho_m}{1 - \rho_m}$$

mittlere Warteschlangenlänge *vor* der Station m: $(m = 1, 2, \ldots, M)$

$$L_m^{\mathrm{Q}} = \frac{\rho_m^2}{1 - \rho_m}$$

[Spezialfall:]

- exponentialverteilte Zwischenankunftszeiten (Poisson-Ankunftsprozess)
- exponentialverteilte Bearbeitungszeiten, d. h. Stationszeiten

Auslastung der Station m: $(m = 1, 2, \ldots, M)$

$$\rho_m = \frac{\lambda_m}{\mu_m} = \lambda_m \cdot b_m$$

Zustandswahrscheinlichkeiten für die Station m: $(m = 1, 2, \ldots, M)$

$$P_{n_m} = (1 - \rho_m) \cdot \rho_m^{n_m} \qquad (n_m = 0, 1, 2, \ldots)$$

mittlerer Bestand *an* der Station m: $(m = 1, 2, \ldots, M)$

$$L_m = \frac{\rho_m}{1 - \rho_m}$$

mittlere Durchlaufzeit an der Station m: $(m = 1, 2, \ldots, M)$

$$W_m = \frac{b_m}{1 - \rho_m}$$

Beispiel: Leistungsanalyse bei M/M/1-Subsystemen

- Reihenproduktion mit insgesamt $M = 5$ Stationen
- Ankunftsrate vor der ersten Station: $\lambda = 0.08$ Stück pro Minute
- mittlere Bearbeitungszeit an allen Stationen: $b_m = 10$ Minuten

Ankunftsrate an Station m:

$$\lambda_1 = \lambda = 0.08,\ \lambda_m = \min\{\lambda_{m-1}, \mu_{m-1}\} = 0.08 \qquad (m = 2, \ldots, 5)$$

Stationsauslastung $U_m = \lambda_m \cdot b_m = 0.08 \cdot 10 = 0.8$

Wahrscheinlichkeitsverteilung für den Bestand an den Stationen

$$P_0 = 0.2,\ P_1 = 0.16,\ P_2 = 0.128,\ P_3 = 0.1024,\ P_4 = 0.0819, \ldots, P_{18} = 0.0036, \ldots$$

Bestand $L_m = \dfrac{\rho_m}{1-\rho_m} = \dfrac{0.8}{0.2} = 4,\ L = \sum_{i=1}^{5} L_m = 20$

Durchlaufzeit $W_m = \dfrac{b_m}{1-\rho_m} = \dfrac{10}{0.2} = 50,\ W = \sum_{i=1}^{5} W_m = 250$

Produktionsrate $X = \min\{\lambda, \mu_1, \mu_2, \mu_3, \mu_4, \mu_5\} = 0.08$