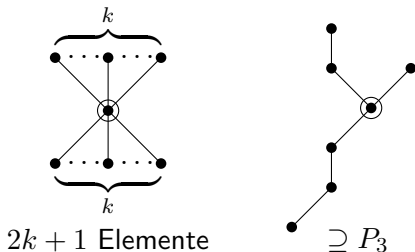


## 4. Schönhage/Paterson/Pippenger-Median-Algorithmus

### Definition 81

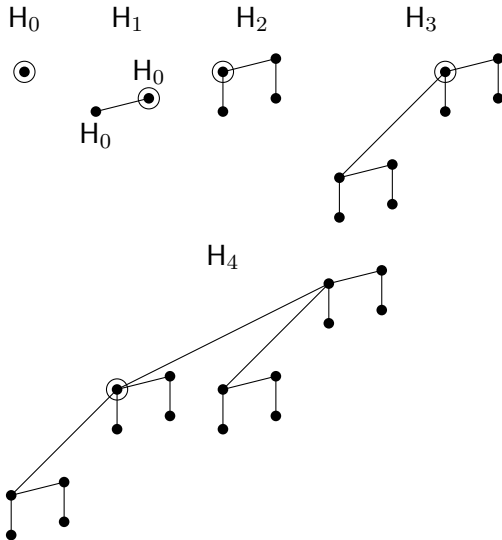
Sei  $k \in \mathbb{N} \setminus \{0\}$ .  $P_k$  sei die folgende partielle Ordnung:



Wir betrachten nun spezielle Binomialbäume mit „Zentrum“ (um Ordnungsinformation widerzuspiegeln).

## Definition 82

- 1 Der Baum  $H_0$  besteht aus einem Knoten, und dieser ist auch das Zentrum.
- 2  $H_{2h}$  ( $h > 0$ ) besteht aus zwei  $H_{2h-1}$ , deren Zentren durch eine neue Kante verbunden sind. Das Zentrum des  $H_{2h}$  ist das kleinere der beiden Zentren der  $H_{2h-1}$ .
- 3  $H_{2h+1}$  ( $h \geq 0$ ) besteht aus zwei  $H_{2h}$ , deren Zentren durch eine neue Kante verbunden sind, sein Zentrum ist das größere dieser beiden Zentren.



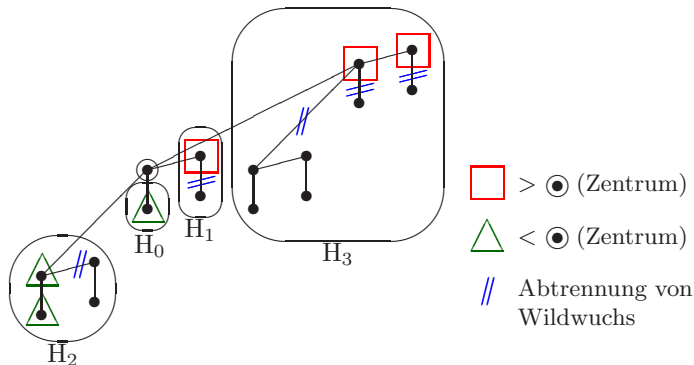
## Lemma 83 (Zerlegungslemma)

- a)  $H_h$  hat  $2^h$  Knoten, es werden  $2^h - 1$  Vergleiche benötigt, um  $H_h$  zu konstruieren.
- b)  $H_{2h}$  kann zerlegt werden in
- sein Zentrum
  - eine Menge  $\{H_1, H_3, \dots, H_{2h-1}\}$  von disjunkten Teilbäumen, deren Zentren alle größer sind als das Zentrum von  $H_{2h}$ .
  - eine Menge  $\{H_0, H_2, H_4, \dots, H_{2h-2}\}$  von disjunkten Teilbäumen mit Zentren kleiner als das von  $H_{2h}$ .

## Lemma 83 (Zerlegungslemma)

- c)  $H_{2h+1}$  kann so zerlegt werden, dass die Zusammenhangskomponente des Zentrums genau  $2^h$  Knoten  $\geq$  dem Zentrum enthält, indem höchstens  $2^{h+1} - 1$  Kanten entfernt werden.  
 $H_{2h}$  kann so zerlegt werden, dass die Zusammenhangskomponente des Zentrums genau  $2^h$  Knoten enthält, die alle  $\leq$  dem Zentrum sind, indem höchstens  $2^h - 1$  Kanten entfernt werden.
- d) Falls  $k \leq 2^h - 1$ , dann kann  $H_{2h}$  so zerlegt werden, dass die Zusammenhangskomponente des Zentrums genau  $2k + 1$  Elemente enthält, von denen  $k$  größer und  $k$  kleiner als das Zentrum sind ( $\Rightarrow P_k$ ).  
Dazu genügt es, höchstens  $3k + 2h$  Kanten zu entfernen. Die restlichen Zusammenhangskomponenten sind wieder  $H_i$ 's.

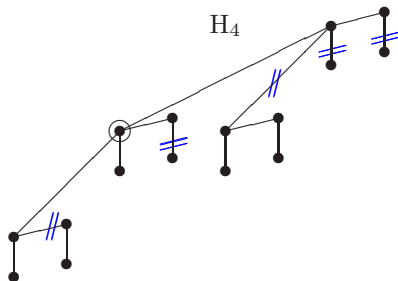
# Zerlegungslemma



**Bemerkung:** Bei jedem Konstruktionsschritt wird ein Vergleich durchgeführt, um zu bestimmen, welcher der beiden Teilbäume das kleinere Zentrum hat. Im Algorithmus von Schönhage, Paterson und Pippenger werden aus Teilstücken  $H_r$  größere Bäume  $H_{r+1}$  zusammengebaut, wodurch schrittweise eine partielle Ordnung auf den Eingabewerten bestimmt wird. Wurde ein Baum  $H_{2^h}$  hinreichender Größe hergestellt, so wird er durch Zerlegung in einen Baum umgewandelt, der nur noch sein altes Zentrum sowie  $k$  darüberliegende und  $k$  darunterliegende Elemente enthält, wobei  $k \leq 2^h - 1$ .

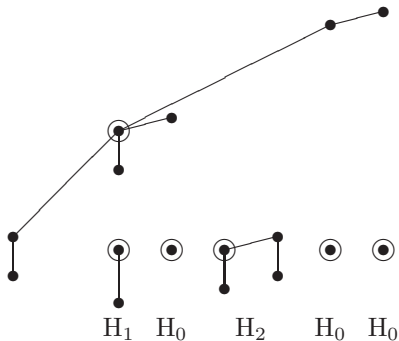
## Beispiel 84

In diesem Beispiel wollen wir  $H_4$  zerlegen und wählen  $k = 3$ :

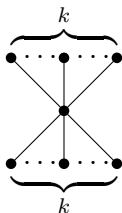




Um einen  $H_4$  derart zu zerlegen, müssen wir 5 Kanten aufbrechen.  
 Dabei werden drei  $H_0$ , ein  $H_1$  sowie ein  $H_2$  abgespalten.



Übrig bleibt die gewünschte Struktur mit  $k$  Knoten über dem Zentrum und  $k$  unter dem Zentrum, wodurch eine partielle Ordnung auf  $2k + 1$  Eingabewerten bestimmt wurde:



$2k + 1$  Elemente

Die bei der Zerlegung angefallenen Reststücke werden beim Aufbau weiterer Bäume benutzt. So geht das bereits angesammelte Wissen über die Ordnung der Elemente nicht verloren.

## Beweis des Zerlegungslemmas:

Wir beginnen mit Teil a).

### Lemma 85

$H_r$  hat  $2^r$  Knoten, es werden  $2^r - 1$  Vergleiche benötigt, um  $H_r$  aufzubauen.

### Beweis:

In jedem der  $r$  Konstruktionsschritte wird die Anzahl der Knoten verdoppelt. Da wir mit einem Knoten beginnen, hat  $H_r$  folglich  $2^r$  Knoten. Die Anzahl der notwendigen Vergleiche  $C_r$  unterliegt folgender Rekursionsgleichung ( $r \geq 1$ ):

$$C_r = 1 + 2C_{r-1} \text{ und } C_0 = 0.$$

Damit folgt sofort  $C_r = 2^r - 1$ . □

## Beweis von b):

### Lemma 86

$H_r$  kann in folgende disjunkte Bereiche unterteilt werden:

- sein Zentrum,
- eine Reihe  $H_1, H_3, \dots, H_{r-1}$  (falls  $r$  gerade) bzw.  $H_1, H_3, \dots, H_{r-2}$  (falls  $r$  ungerade) von Unterbäumen, deren Zentren alle über dem von  $H_r$  liegen,
- eine Reihe  $H_0, H_2, \dots, H_{r-2}$  (falls  $r$  gerade) bzw.  $H_0, H_2, \dots, H_{r-1}$  (falls  $r$  ungerade) von Unterbäumen, deren Zentren alle unter dem von  $H_r$  liegen.

## Beweis von b):

Beweis:

Durch Induktion über  $r$ .

**Induktionsanfang:** für  $H_0$  gilt die Behauptung.

**Induktionsannahme:** die Behauptung gelte für  $H_{r-1}$ .

## Beweis von b):

### Beweis:

- ① Sei  $r = 2h, h > 0$ .

$H_{2h}$  besteht aus zwei  $H_{2h-1}$ , wobei das kleinere der beiden alten Zentren das neue Zentrum  $z$  bildet. Wende auf den  $H_{2h-1}$ , der  $z$  enthält, die Induktionsannahme an. Wir können diesen Unterbaum also in  $z$  sowie  $H_1, H_3, \dots, H_{2h-3}$  (Zentren **über**  $z$ ) und  $H_0, H_2, \dots, H_{2h-2}$  (Zentren **unter**  $z$ ) partitionieren. Zusammen mit dem  $H_{2h-1}$ , dessen Zentrum über  $z$  liegt, ergibt sich die Induktionsbehauptung für  $H_r = H_{2h}$ .

## Beweis von b):

### Beweis:

- ② Sei  $r = 2h + 1$ ,  $h \geq 0$ .  $H_{2h+1}$  besteht aus zwei  $H_{2h}$ , wobei das größere der beiden alten Zentren das neue Zentrum  $z$  bildet. Wende auf den  $H_{2h}$ , der  $z$  enthält, die Induktionsannahme an. Wir können diesen Unterbaum also in  $z$  sowie  $H_1, H_3, \dots, H_{2h-1}$  (Zentren **über**  $z$ ) und  $H_0, H_2, \dots, H_{2h-2}$  (Zentren **unter**  $z$ ) partitionieren. Zusammen mit dem  $H_{2h}$ , dessen Zentrum unter  $z$  liegt, ergibt sich die Induktionsbehauptung für  $H_r = H_{2h+1}$ .



## Beweis von c):

Wir bezeichnen im Folgenden mit  $H_{2h}^-$  den Baum, der entsteht, wenn wir  $H_{2h}$  so zerlegen, dass der verbleibende Rest nur mehr Elemente unterhalb des Zentrums (und das Zentrum selbst) enthält. Mit  $H_{2h+1}^+$  bezeichnen wir den Baum, der entsteht, wenn wir  $H_{2h+1}$  so zerlegen, dass der verbleibende Rest nur mehr Elemente über dem Zentrum und dieses selbst enthält.

### Lemma 87

*$H_{2h}^-$  und  $H_{2h+1}^+$  haben jeweils  $2^h$  Knoten. Bei der Herstellung aus  $H_{2h}$  bzw.  $H_{2h+1}$  werden  $2^h - 1$  bzw.  $2^{h+1} - 1$  Kanten aufgebrochen. Die wegfallenden Teile haben die Form  $H_s$ ,  $s < 2h$  bzw.  $s < 2h + 1$ .*



## Beweis:

Durch Induktion über  $r$ .

**Induktionsanfang:** für  $H_0$  und  $H_1$  gilt die Behauptung.

**Induktionsannahme:** die Behauptung gilt für alle  $H_p$ ,  $p < r$ .

- 1 Sei  $r = 2h$ ,  $h > 0$ . Wir betrachten die Partitionierung von  $H_{2h}$  mit Zentrum  $z$  wie in Lemma 86. Die Unterbäume  $H_1, H_3, \dots, H_{2h-1}$  haben ihre Zentren oberhalb von  $z$ . Wir trennen sie von  $H_{2h}$ , indem wir  $h$  Kanten aufbrechen. Die abgetrennten Teile haben offensichtlich die Form  $H_s$ ,  $s < 2h$ . Bei den Unterbäumen  $H_0, H_2, \dots, H_{2h-2}$ , mit Zentren unterhalb von  $z$ , wenden wir jeweils die Induktionsannahme an, d.h. wir erzeugen  $H_0^-, H_2^-, \dots, H_{2h-2}^-$ . Als Ergebnis erhalten wir  $H_{2h}^-$ .

## Beweis (Forts.):

Damit gilt für die Zahl der aufzubrechenden Kanten  $K^-(2h)$  zur Herstellung von  $H_{2h}^-$ :

$$K^-(2h) = h + \sum_{i=0}^{h-1} K^-(2i) \stackrel{I.A.}{=} h + \sum_{i=0}^{h-1} (2^i - 1) = \sum_{i=0}^{h-1} 2^i = 2^h - 1.$$

Für die Zahl  $E^-(2h)$  der Elemente in  $H_{2h}^-$  gilt:

$$E^-(2h) = 1 + \sum_{i=0}^{h-1} E^-(2i) \stackrel{I.A.}{=} 1 + \sum_{i=0}^{h-1} 2^i = 1 + \underbrace{\sum_{i=1}^h 2^{i-1}}_{2^h - 1} = 2^h.$$

## Beweis (Forts.):

- ② Sei  $r = 2h + 1$ ,  $h > 0$ . Wir betrachten die Partitionierung von  $H_{2h+1}$  mit Zentrum  $z$  wie in Lemma 86. Die Unterbäume  $H_0, H_2, \dots, H_{2h}$  haben ihre Zentren unterhalb von  $z$ . Wir trennen sie von  $H_{2h+1}$ , indem wir  $h + 1$  Kanten aufbrechen. Die abgetrennten Teile haben offensichtlich die Form  $H_s$ ,  $s < 2h + 1$ . Bei den Unterbäumen  $H_1, H_3, \dots, H_{2h-1}$ , mit Zentren oberhalb von  $z$ , wenden wir jeweils die Induktionsannahme an, d.h. wir erzeugen  $H_1^+, H_3^+, \dots, H_{2h-1}^+$ . Als Ergebnis erhalten wir  $H_{2h+1}^+$ . Damit gilt für die Zahl der aufzubrechenden Kanten  $K^+(2h + 1)$  zur Herstellung von  $H_{2h+1}^+$ :

## Beweis (Forts.):

$$\begin{aligned}K^+(2h+1) &= h+1 + \sum_{i=1}^h K^+(2(i-1)+1) \\ &\stackrel{I.A.}{=} h+1 + \sum_{i=1}^h (2^i - 1) = 1 + \sum_{i=1}^h 2^i \\ &= 1 + \underbrace{\sum_{i=1}^{h+1} 2^{i-1}}_{2^{h+1}-1} - 1 = 2^{h+1} - 1.\end{aligned}$$

Für die Zahl  $E^+(2h+1)$  der Elemente in  $H_{2h+1}^+$  gilt:

$$E^+(2h+1) = 1 + \sum_{i=1}^h E^+(2(i-1)+1) \stackrel{I.A.}{=} 1 + \underbrace{\sum_{i=1}^h 2^{i-1}}_{2^h-1} = 2^h.$$



## Beweis von d):

### Lemma 88

*Falls  $k \leq 2^h - 1$ , dann kann  $H_{2h}$  so zerlegt werden, dass die Komponente des Zentrums genau  $2k + 1$  Elemente enthält,  $k$  davon über und  $k$  unter dem Zentrum. Dazu müssen  $\leq 3k + 2h$  Kanten entfernt werden. Die entfernten Teile sind von der Form  $H_s$ ,  $s < 2h$ .*

### Beweis:

Betrachte die Binärdarstellung von  $k = k_0 2^0 + k_1 2^1 + \dots + k_{h-1} 2^{h-1}$  und die Partitionierung von  $H_{2h}$  mit Zentrum  $z$  wie in Lemma 86.

## Beweis (Forts.):

Für jedes  $i$  mit  $k_i = 1$ , betrachte  $H_{2^{i+1}}$  aus der Sequenz  $H_1, H_3, \dots, H_{2^{h-1}}$  von Unterbäumen, deren Zentren oberhalb von  $z$  liegen, und schneide alle Elemente aus  $H_{2^{i+1}}$ , die kleiner als sein Zentrum sind (bilde also  $H_{2^{i+1}}^+$ ). Dazu müssen höchstens  $2k$  Kanten aufgebrochen werden, denn jedes  $k_i = 1$  steht für  $2^i$  in  $k$ , kostet aber nach Lemma 87  $K^+(2i+1) = 2^{i+1} - 1$  Kanten, also:

$$\sum_{i=0}^{h-1} k_i K^+(2i+1) \leq 2k.$$

Für jedes  $i$  mit  $k_i = 0$ , schneide  $H_{2^{i+1}}$  ganz weg. Dabei werden  $\leq h$  Kanten aufgebrochen. Genau  $k$  Elemente oberhalb  $z$  bleiben zurück, da jedes  $k_i = 1$  für  $2^i$  in  $k$  steht, und ein  $H_{2^{i+1}}^+$  genau  $E^+(2i+1) = 2^i$  Elemente enthält, also:

$$\sum_{i=0}^{h-1} k_i E^+(2i+1) = k.$$

## Beweis (Forts.):

Für jedes  $i$  mit  $k_i = 1$ , betrachte  $H_{2^i}$  aus der Sequenz  $H_0, H_2, \dots, H_{2^{h-2}}$  von Unterbäumen, deren Zentren unterhalb von  $z$  liegen, und schneide alle Elemente aus  $H_{2^i}$ , die größer als sein Zentrum sind (bilde also  $H_{2^i}^-$ ). Dazu müssen höchstens  $k - 1$  Kanten aufgebrochen werden, denn jedes  $k_i = 1$  steht für  $2^i$  in  $k$  und kostet uns nach Lemma 87  $K^-(2^i) = 2^i - 1$  Kanten, also:

$$\sum_{i=0}^{h-1} k_i(2^i - 1) \leq k - 1.$$

Für jedes  $i$  mit  $k_i = 0$ , schneide  $H_{2^i}$  ganz weg. Dabei werden höchstens  $h$  Kanten aufgebrochen. Genau  $k$  Elemente unterhalb von  $z$  bleiben zurück, da jedes  $k_i = 1$  für  $2^i$  in  $k$  steht, und ein  $H_{2^i}^-$  genau  $E^-(2^i) = 2^i$  Elemente enthält, also:

$$\sum_{i=0}^{h-1} k_i E^-(2^i) = k.$$

## Beweis (Forts.):

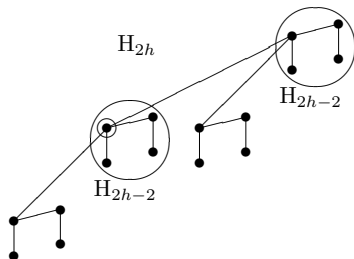
Damit ergibt sich für die Gesamtanzahl aufzubrechender Kanten eine obere Schranke von  $3k + 2h$ . Lemma 87 liefert uns darüber hinaus die gewünschte Aussage über die Form der abgetrennten Teile. □



## Beweis von d):

Betrachte  $H_{2h}$ .

- „größer“:  $H_{2h-1}, H_{2h-3}, \dots, H_1$
- „kleiner“:  $H_{2h-2}, H_{2h-4}, \dots, H_0$



$U(h) :=$  Anzahl der Elemente in  $H_{2h} \geq$  Zentrum:

$$U(h) = 2U(h-1) = 2^h; \quad U(0) = 1$$

$D(h) :=$  Anzahl der Elemente in  $H_{2h} \leq$  Zentrum:

$$D(h) = 2D(h-1) = 2^h; \quad D(0) = 1$$

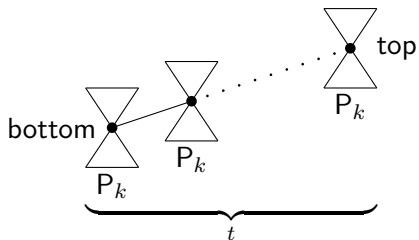
## Beweis von d):

Anzahl der Kanten, die entfernt werden müssen:

$$\left. \begin{aligned} C_u(h) &\leq 2 + 2C_u(h-1) \\ &= 2 + 4 + 2^3 + \dots + 2^h \\ &= 2^{h+1} - 2 \\ C_d(h) &\leq 1 + 2C_d(h-1) \\ &= 2^h - 1 \end{aligned} \right\} C(h) \leq 2^{h+1} - 2 + 2^h - 1 < 3 \cdot 2^h$$

Damit ist der Beweis des Zerlegungslemmas beendet. □

Kette von  $P_k$ 's:



Gesamtzahl der Elemente:

$n$

$t(2k + 1)$  in den  $P_k$ 's

$r = n - t(2k + 1)$  Rest

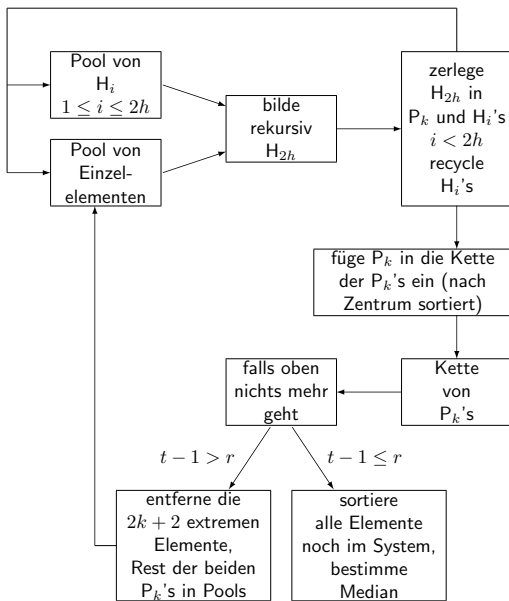
Wenn  $r < t - 1$ , dann wissen wir, dass **top** größer ist als

$$k + (t - 1)(k + 1) > k + (k + 1) \left( \frac{n + 1}{2k + 2} - 1 \right) = \frac{n - 1}{2}$$

$\Rightarrow$  top  $>$  Median (Entsprechendes gilt für das Element **bottom**)

Setze

$$k := \left\lfloor n^{\frac{1}{4}} \right\rfloor$$
$$h \text{ sdg. } 2^{h-1} \leq k < 2^h$$



Definiere  $r :=$  Anzahl der noch im  $H_{2h}$ -Produktionsprozess steckenden Elemente (für jedes  $i < 2h$  höchstens ein  $H_i$ , daher

$$r \leq \sum_{i=0}^{2h-1} 2^i = 2^{2h} - 1).$$

$R :=$  Anzahl der im letzten Schritt zu sortierenden Elemente. Es gilt:  $t \leq r + 1$ , und damit

$$R = t(2k + 1) + r \leq 2^{2h}(2k + 1) + 2^{2h} - 1.$$

$m :=$  Gesamtzahl der im Algorithmus produzierten  $P_k$ 's.

$$m = t + 2 \frac{n - R}{2(k + 1)} = t + \frac{n - R}{k + 1}.$$

Also:  $r$  und  $t$  sind  $\mathcal{O}(n^{\frac{1}{2}})$ ,  $R$  ist  $\mathcal{O}(n^{\frac{3}{4}})$ .

Gesamtzahl der vom Algorithmus durchgeführten Vergleiche =

- ① Anzahl der Kanten in allen  $P_k$ 's
- ② + Anzahl der Kanten, die gelöscht werden, um die  $P_k$ 's zu formen
- ③ + Anzahl der Kanten, die zum Schluss in übriggebliebenen  $H_i$ 's,  $i < 2h$ , stecken
- ④ + Anzahl der Vergleiche, um jedes Zentrum der  $P_k$ 's in die (sortierte) Kette einzufügen
- ⑤ + Anzahl der Vergleiche, um die zum Schluss übriggebliebenen  $R$  Elemente zu sortieren

$$\leq \left( \frac{n-R}{k+1} + t \right) \left[ \underbrace{2k}_1 + \underbrace{3k+2h}_2 + \underbrace{\log \frac{n}{2k+1}}_4 \right] + \underbrace{R \log R}_5 + \underbrace{r}_3 .$$

Mit  $k = \left\lceil n^{\frac{1}{4}} \right\rceil$ ,  $h$  so, dass  $2^{h-1} \leq k < 2^h$ , ergibt sich damit

$$r = \mathcal{O}(k^2)$$

$$t = \mathcal{O}(k^2) \text{ zum Schluss}$$

$$R = \mathcal{O}(k^3), \text{ und damit die}$$

$$\text{Anzahl der Vergleiche} = T(n) \leq 5n + o(n).$$

Verbesserte Version (besseres Zurechtschneiden, bessere Verwertung der Reste):

$$T(n) = 3n + o(n)$$



Bester bekannter Algorithmus (von Dor/Zwick):

$$2,95n + o(n)$$

## Literatur:



Arnold Schönhage, Michael Paterson, Nicholas Pippenger:  
*Finding the median*  
J. Comput. Syst. Sci. **13**, pp. 184–199 (1976)



Dorit Dor, Uri Zwick:  
*Selecting the median*  
SIAM J. Comput. **28**(5), pp. 1722–1758 (1999)