

5. DIVIDE AND CONQUER I

- ▶ *mergesort*
- ▶ *counting inversions*
- ▶ *randomized quicksort*
- ▶ *median and selection*
- ▶ *closest pair of points*

Lecture slides by Kevin Wayne

Copyright © 2005 Pearson–Addison Wesley

<http://www.cs.princeton.edu/~wayne/kleinberg-tardos>

Divide-and-conquer paradigm

Divide-and-conquer.

- Divide up problem into several subproblems (of the same kind).
- Solve (conquer) each subproblem recursively.
- Combine solutions to subproblems into overall solution.

Most common usage.

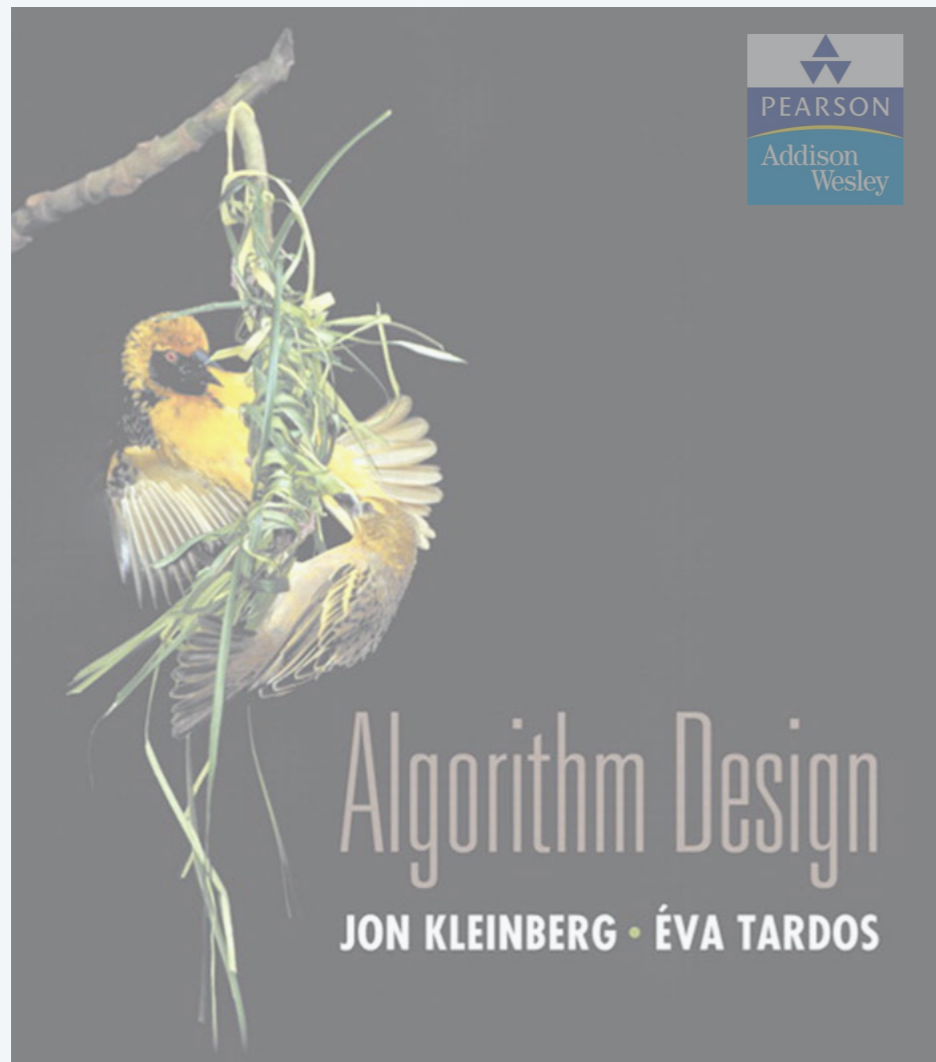
- Divide problem of size n into **two** subproblems of size $n/2$. ← $O(n)$ time
- Solve (conquer) two subproblems recursively.
- Combine two solutions into overall solution. ← $O(n)$ time

Consequence.

- Brute force: $\Theta(n^2)$.
- Divide-and-conquer: $O(n \log n)$.



attributed to Julius Caesar



SECTIONS 5.1–5.2

5. DIVIDE AND CONQUER

- ▶ *mergesort*
- ▶ *counting inversions*
- ▶ *randomized quicksort*
- ▶ *median and selection*
- ▶ *closest pair of points*

Sorting problem

Problem. Given a list L of n elements from a totally ordered universe, rearrange them in ascending order.



The image shows a music player interface. At the top, there are several album covers displayed in a row. The central cover is for Bruce Springsteen's "Born In The U.S.A.". Below the covers is a progress bar and a play button. Below the player is a list of songs with columns for Name, Artist, Time, and Album. The song "Dancing In The Dark" by Bruce Springsteen is highlighted in blue.

	Name	Artist	Time	Album
12	<input checked="" type="checkbox"/> Let It Be	The Beatles	4:03	Let It Be
13	<input checked="" type="checkbox"/> Take My Breath Away	BERLIN	4:13	Top Gun – Soundtrack
14	<input checked="" type="checkbox"/> Circle Of Friends	Better Than Ezra	3:27	Empire Records
15	<input checked="" type="checkbox"/> Dancing With Myself	Billy Idol	4:43	Don't Stop
16	<input checked="" type="checkbox"/> Rebel Yell	Billy Idol	4:49	Rebel Yell
17	<input checked="" type="checkbox"/> Piano Man	Billy Joel	5:36	Greatest Hits Vol. 1
18	<input checked="" type="checkbox"/> Pressure	Billy Joel	3:16	Greatest Hits, Vol. II (1978 – 1985) (Disc 2)
19	<input checked="" type="checkbox"/> The Longest Time	Billy Joel	3:36	Greatest Hits, Vol. II (1978 – 1985) (Disc 2)
20	<input checked="" type="checkbox"/> Atomic	Blondie	3:50	Atomic: The Very Best Of Blondie
21	<input checked="" type="checkbox"/> Sunday Girl	Blondie	3:15	Atomic: The Very Best Of Blondie
22	<input checked="" type="checkbox"/> Call Me	Blondie	3:33	Atomic: The Very Best Of Blondie
23	<input checked="" type="checkbox"/> Dreaming	Blondie	3:06	Atomic: The Very Best Of Blondie
24	<input checked="" type="checkbox"/> Hurricane	Bob Dylan	8:32	Desire
25	<input checked="" type="checkbox"/> The Times They Are A-Changin'	Bob Dylan	3:17	Greatest Hits
26	<input checked="" type="checkbox"/> Livin' On A Prayer	Bon Jovi	4:11	Cross Road
27	<input checked="" type="checkbox"/> Beds Of Roses	Bon Jovi	6:35	Cross Road
28	<input checked="" type="checkbox"/> Runaway	Bon Jovi	3:53	Cross Road
29	<input checked="" type="checkbox"/> Rasputin (Extended Mix)	Boney M	5:50	Greatest Hits
30	<input checked="" type="checkbox"/> Have You Ever Seen The Rain	Bonnie Tyler	4:10	Faster Than The Speed Of Night
31	<input checked="" type="checkbox"/> Total Eclipse Of The Heart	Bonnie Tyler	7:02	Faster Than The Speed Of Night
32	<input checked="" type="checkbox"/> Straight From The Heart	Bonnie Tyler	3:41	Faster Than The Speed Of Night
33	<input checked="" type="checkbox"/> Holding Out For A Hero	Bonny Tyler	5:49	Meat Loaf And Friends
34	<input checked="" type="checkbox"/> Dancing In The Dark	Bruce Springsteen	4:05	Born In The U.S.A.
35	<input checked="" type="checkbox"/> Thunder Road	Bruce Springsteen	4:51	Born To Run
36	<input checked="" type="checkbox"/> Born To Run	Bruce Springsteen	4:30	Born To Run
37	<input checked="" type="checkbox"/> Jungleland	Bruce Springsteen	9:34	Born To Run
38	<input checked="" type="checkbox"/> Turtl Turtl Turtl (To Everything)	The Buds	3:57	Forest Gump The Soundtrack (Disc 2)

Sorting applications

Obvious applications.

- Organize an MP3 library.
- Display Google PageRank results.
- List RSS news items in reverse chronological order.

Some problems become easier once elements are sorted.

- Identify statistical outliers.
- Binary search in a database.
- Remove duplicates in a mailing list.

Non-obvious applications.

- Convex hull.
- Closest pair of points.
- Interval scheduling / interval partitioning.
- Scheduling to minimize maximum lateness.
- Minimum spanning trees (Kruskal's algorithm).
- ...

Mergesort

- Recursively sort left half.
- Recursively sort right half.
- Merge two halves to make sorted whole.

input

A	L	G	O	R	I	T	H	M	S
---	---	---	---	---	---	---	---	---	---

sort left half

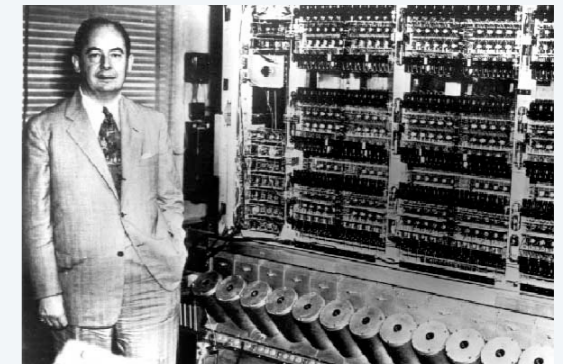
A	G	L	O	R	I	T	H	M	S
---	---	---	---	---	---	---	---	---	---

sort right half

A	G	L	O	R	H	I	M	S	T
---	---	---	---	---	---	---	---	---	---

merge results

A	G	H	I	L	M	O	R	S	T
---	---	---	---	---	---	---	---	---	---



**First Draft
of a
Report on the
EDVAC**
John von Neumann

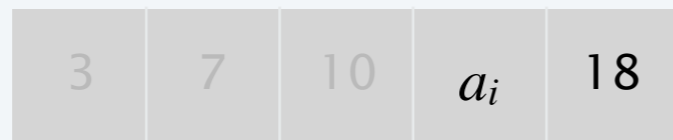
Merging

Goal. Combine two sorted lists A and B into a sorted whole C .

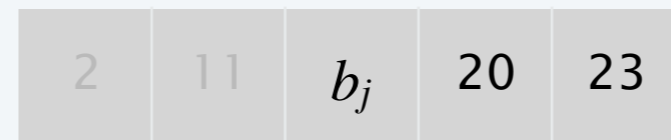


- Scan A and B from left to right.
- Compare a_i and b_j .
- If $a_i \leq b_j$, append a_i to C (no larger than any remaining element in B).
- If $a_i > b_j$, append b_j to C (smaller than every remaining element in A).

sorted list A



sorted list B



merge to form sorted list C



Mergesort implementation

Input. List L of n elements from a totally ordered universe.

Output. The n elements in ascending order.

MERGE-SORT(L)

IF (list L has one element)

RETURN L .

Divide the list into two halves A and B .

$A \leftarrow$ **MERGE-SORT**(A). $\longleftarrow T(n/2)$

$B \leftarrow$ **MERGE-SORT**(B). $\longleftarrow T(n/2)$

$L \leftarrow$ **MERGE**(A, B). $\longleftarrow \Theta(n)$

RETURN L .

A useful recurrence relation

Def. $T(n)$ = max number of compares to mergesort a list of length n .

Recurrence.

$$T(n) \leq \begin{cases} 0 & \text{if } n = 1 \\ T(\lfloor n/2 \rfloor) + T(\lceil n/2 \rceil) + n & \text{if } n > 1 \end{cases}$$

between $\lfloor n/2 \rfloor$ and $n - 1$ compares

Solution. $T(n)$ is $O(n \log_2 n)$.

Assorted proofs. We describe several ways to solve this recurrence.

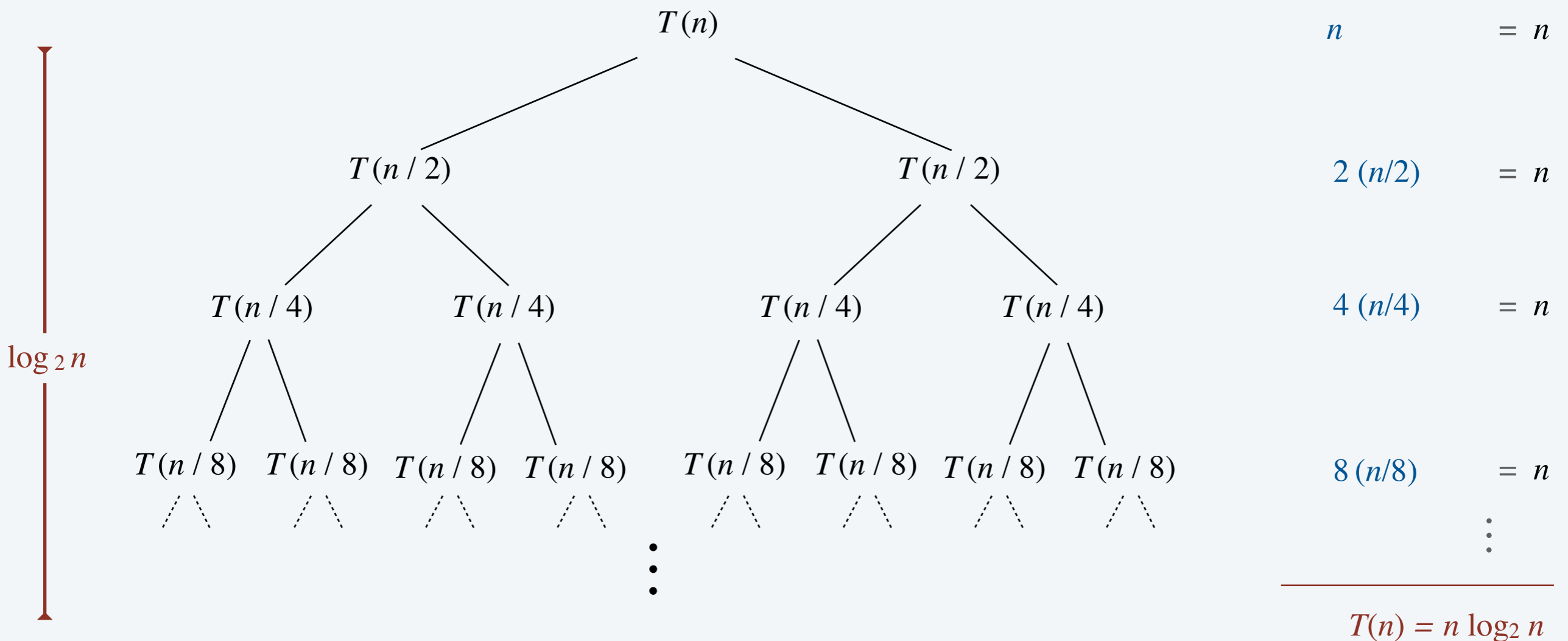
Initially we assume n is a power of 2 and replace \leq with $=$ in the recurrence.

Divide-and-conquer recurrence: recursion tree

Proposition. If $T(n)$ satisfies the following recurrence, then $T(n) = n \log_2 n$.

$$T(n) = \begin{cases} 0 & \text{if } n = 1 \\ 2T(n/2) + n & \text{if } n > 1 \end{cases}$$

assuming n
is a power of 2



Proof by induction

Proposition. If $T(n)$ satisfies the following recurrence, then $T(n) = n \log_2 n$.

$$T(n) = \begin{cases} 0 & \text{if } n = 1 \\ 2T(n/2) + n & \text{if } n > 1 \end{cases}$$

assuming n
is a power of 2

Pf. [by induction on n]

- Base case: when $n = 1$, $T(1) = 0 = n \log_2 n$.
- Inductive hypothesis: assume $T(n) = n \log_2 n$.
- Goal: show that $T(2n) = 2n \log_2 (2n)$.

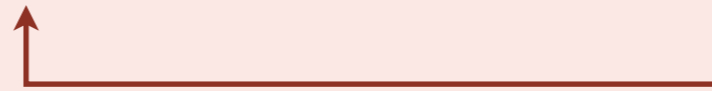
$$T(2n) \stackrel{\text{recurrence}}{=} 2T(n) + 2n$$

$$\begin{aligned} \text{inductive hypothesis} \longrightarrow &= 2n \log_2 n + 2n \\ &= 2n (\log_2 (2n) - 1) + 2n \\ &= 2n \log_2 (2n). \quad \blacksquare \end{aligned}$$



Which is the exact solution of the following recurrence?

$$T(n) = \begin{cases} 0 & \text{if } n = 1 \\ T(\lfloor n/2 \rfloor) + T(\lceil n/2 \rceil) + n - 1 & \text{if } n > 1 \end{cases}$$



no longer assuming n
is a power of 2

- A. $T(n) = n \lfloor \log_2 n \rfloor$
- B. $T(n) = n \lceil \log_2 n \rceil$
- C. $T(n) = n \lfloor \log_2 n \rfloor + 2^{\lfloor \log_2 n \rfloor} - 1$
- D. $T(n) = n \lceil \log_2 n \rceil - 2^{\lceil \log_2 n \rceil} + 1$
- E. Not even Knuth knows.

Analysis of mergesort recurrence

Proposition. If $T(n)$ satisfies the following recurrence, then $T(n) \leq n \lceil \log_2 n \rceil$.

$$T(n) \leq \begin{cases} 0 & \text{if } n = 1 \\ T(\lfloor n/2 \rfloor) + T(\lceil n/2 \rceil) + n & \text{if } n > 1 \end{cases}$$

no longer assuming n is a power of 2

Pf. [by strong induction on n]

- Base case: $n = 1$.
- Define $n_1 = \lfloor n/2 \rfloor$ and $n_2 = \lceil n/2 \rceil$ and note that $n = n_1 + n_2$.
- Induction step: assume true for $1, 2, \dots, n-1$.

$$T(n) \leq T(n_1) + T(n_2) + n$$

inductive hypothesis \longrightarrow $\leq n_1 \lceil \log_2 n_1 \rceil + n_2 \lceil \log_2 n_2 \rceil + n$

$$\leq n_1 \lceil \log_2 n_2 \rceil + n_2 \lceil \log_2 n_2 \rceil + n$$

$$= n \lceil \log_2 n_2 \rceil + n$$

$$\leq n (\lceil \log_2 n \rceil - 1) + n$$

$$= n \lceil \log_2 n \rceil. \quad \blacksquare$$

$$n_2 = \lceil n/2 \rceil$$

$$\leq \left\lceil 2^{\lceil \log_2 n \rceil} / 2 \right\rceil$$

$$= 2^{\lceil \log_2 n \rceil} / 2$$

$$\log_2 n_2 \leq \lceil \log_2 n \rceil - 1$$

an integer