# Complexity of the Adaptive ShiversSort Algorithm and of its sibling TimSort

Vincent Jugé

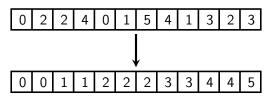
LIGM - Université Paris-Est Marne-la-Vallée, ESIEE, ENPC & CNRS

07/01/2020

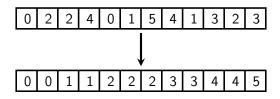
## Contents

- Efficient Merge Sorts
- 2 TimSort
- 3 Adaptive ShiversSort

## Sorting data



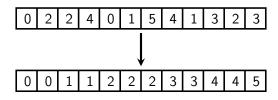
## Sorting data



MergeSort has a worst-case time complexity of  $O(n \log(n))$ 

Can we do better?

## Sorting data



MergeSort has a worst-case time complexity of  $O(n \log(n))$ 

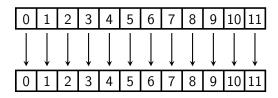
## Can we do better? No!

#### **Proof:**

- There are *n*! possible reorderings
- Each element comparison gives a 1-bit information
- Thus  $\log_2(n!) \sim n \log_2(n)$  tests are required

## Cannot we ever do better?

In some cases, we should...





Chunk your data in non-decreasing runs

5 runs of lengths 4, 3, 1, 2 and 2

											_
0	2	2	4	0	1	5	4	1	3	2	3

- Chunk your data in non-decreasing runs
- ② New parameters: Number of runs  $(\rho)$  and their lengths  $(r_1, \ldots, r_{\rho})$

5 runs of lengths 4, 3, 1, 2 and 2

		_					_				
0	2	2	4	0	1	5	4	1	3	2	3

- Chunk your data in non-decreasing runs
- ② New parameters: Number of runs  $(\rho)$  and their lengths  $(r_1,\ldots,r_{
  ho})$

Run-length entropy: 
$$\mathcal{H} = \sum_{i=1}^{\rho} (r_i/n) \log_2(n/r_i)$$
  
 $\leq \log_2(\rho) \leq \log_2(n)$ 

5 runs of lengths 4, 3, 1, 2 and 2

0	2	2	4	0	1	5	4	1	3	2	3

- Chunk your data in non-decreasing runs
- ② New parameters: Number of runs (
  ho) and their lengths  $(r_1,\ldots,r_
  ho)$

Run-length entropy: 
$$\mathcal{H} = \sum_{i=1}^{\rho} (r_i/n) \log_2(n/r_i)$$
  
 $\leq \log_2(\rho) \leq \log_2(n)$ 

## Theorem [7]

TimSort has a worst-case time complexity of  $\mathcal{O}(n+n\mathcal{H})$ 

5 runs of lengths 4, 3, 1, 2 and 2

0	2	2	4	0	1	5	4	1	3	2	3

- Chunk your data in non-decreasing runs
- ② New parameters: Number of runs (
  ho) and their lengths  $(r_1,\ldots,r_
  ho)$

Run-length entropy: 
$$\mathcal{H} = \sum_{i=1}^{\rho} (r_i/n) \log_2(n/r_i)$$
  
 $\leq \log_2(\rho) \leq \log_2(n)$ 

## Theorem [7]

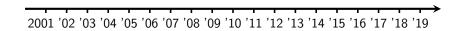
TimSort has a worst-case time complexity of  $\mathcal{O}(n+n\mathcal{H})$ 

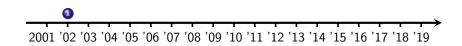
## We cannot do better than $\Omega(n + n\mathcal{H})!^{[4]}$

- Reading the whole input requires a time  $\Omega(n)$
- There are **X** possible reorderings, with  $X \geqslant 2^{1-\rho} \binom{n}{r_1 \dots r_\rho} \geqslant 2^{n \mathcal{H}/2}$

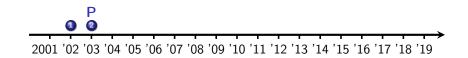
## Contents

- Efficient Merge Sorts
- 2 TimSort
- Adaptive ShiversSort

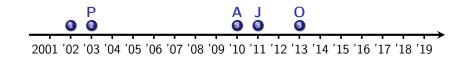




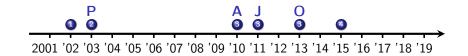
Invented by Tim Peters<sup>[3]</sup>



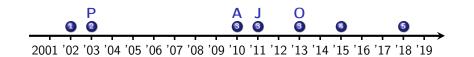
- Invented by Tim Peters<sup>[3]</sup>
- Standard algorithm in Python



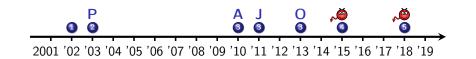
- Invented by Tim Peters<sup>[3]</sup>
- Standard algorithm in Python
- 6 for non-primitive arrays in Android, Java, Octave



- Invented by Tim Peters<sup>[3]</sup>
- Standard algorithm in Python
- for non-primitive arrays in Android, Java, Octave
- lacktriangledown 1st worst-case complexity analysis [6] TimSort works in time  $\mathcal{O}(n \log n)$

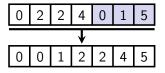


- Invented by Tim Peters<sup>[3]</sup>
- Standard algorithm in Python
- 6 for non-primitive arrays in Android, Java, Octave
- lacktriangledown  $1^{\mathrm{st}}$  worst-case complexity analysis [6] TimSort works in time  $\mathcal{O}(n\log n)$
- **3** Refined worst-case analysis<sup>[7]</sup> TimSort works in time  $\mathcal{O}(n+n\mathcal{H})$



- 1 Invented by Tim Peters<sup>[3]</sup>
- Standard algorithm in Python
- 6 for non-primitive arrays in Android, Java, Octave
- f 0 1st worst-case complexity analysis f [6] TimSort works in time  $m{\mathcal{O}}(n\log n)$
- **3** Refined worst-case analysis<sup>[7]</sup> TimSort works in time  $\mathcal{O}(n+n\mathcal{H})$
- Bugs uncovered in Python & Java implementations<sup>[5,7]</sup>

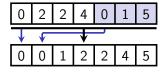
Algorithm based on merging adjacent runs



Algorithm based on merging adjacent runs 

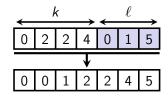
Stable algorithm

 Stable algorithm (good for composite types)



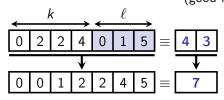
Algorithm based on merging adjacent runs Stable algorithm

(good for composite types)



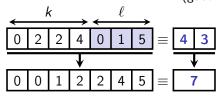
- Run merging algorithm: standard + many optimizations

Algorithm based on merging adjacent runs Stable algorithm (good for composite types)



- Run merging algorithm: standard + many optimizations
- Policy for choosing runs to merge:
  - depends on run lengths only

Algorithm based on merging adjacent runs Stable algorithm (good for composite types)



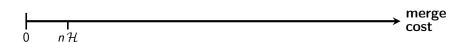
- Run merging algorithm: standard + many optimizations
  - ► time  $\mathcal{O}(k+\ell)$ ► memory  $\mathcal{O}(\min(k,\ell))$ Merge cost:  $k+\ell$
- Policy for choosing runs to merge:
  - depends on run lengths only
- Complexity analysis:
  - Evaluate the total merge cost
  - Forget array values and only work with run lengths

## Some results about merge costs

#### Best-case merge costs:

ullet Every algorithm has a best-case merge cost of at least  $n\,\mathcal{H}^{[4,10]}$ 

Worst-case merge costs:



## Some results about merge costs

#### Best-case merge costs:

• Every algorithm has a best-case merge cost of at least  $n\mathcal{H}^{[4,10]}$ 

#### Worst-case merge costs:

• TimSort has a worst-case merge cost of  $3/2 n \mathcal{H} + \mathcal{O}(n)^{[7,9]}$ 



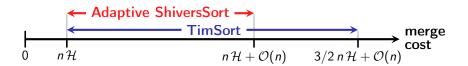
## Some results about merge costs

#### Best-case merge costs:

• Every algorithm has a best-case merge cost of at least  $n\mathcal{H}^{[4,10]}$ 

#### Worst-case merge costs:

- TimSort has a worst-case merge cost of  $3/2 n \mathcal{H} + \mathcal{O}(n)^{[7,9]}$
- Adaptive ShiversSort has a worst-case merge cost of  $n \mathcal{H} + \mathcal{O}(n)^{[10]}$

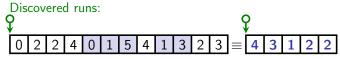


## Contents

- Efficient Merge Sorts
- 2 TimSort
- Adaptive ShiversSort

**STACK** 

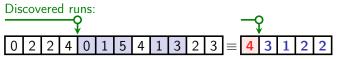
- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack
  - 2 merge the top 1<sup>st</sup> and 2<sup>nd</sup> runs
  - merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



STACK

- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack

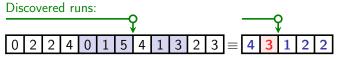
  - 3 merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



4 STACK

- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack

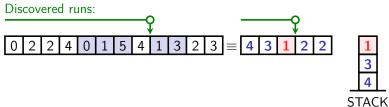
  - 3 merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



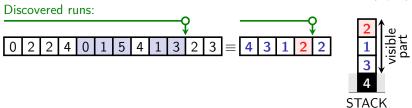


- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack

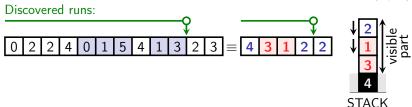
  - 3 merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack
  - merge the top 1<sup>st</sup> and 2<sup>nd</sup> runs
  - 3 merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs

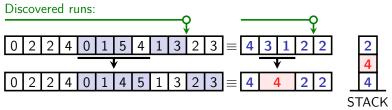


- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack
  - 2 merge the top 1<sup>st</sup> and 2<sup>nd</sup> runs
  - merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



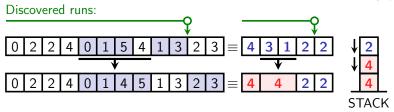
- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack

  - 3 merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs

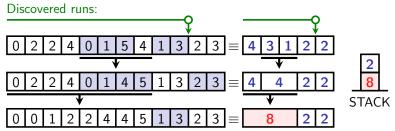


- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack

  - 3 merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs

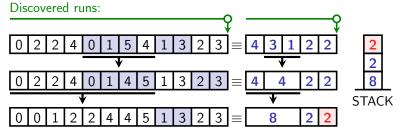


- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack
  - 2 merge the top 1<sup>st</sup> and 2<sup>nd</sup> runs
  - 3 merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



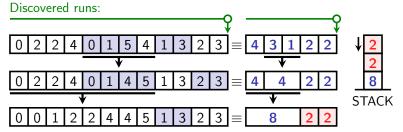
- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack

  - 3 merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



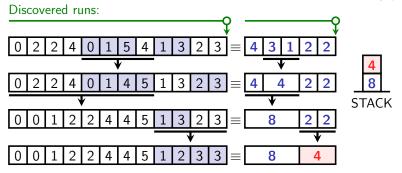
- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack

  - merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



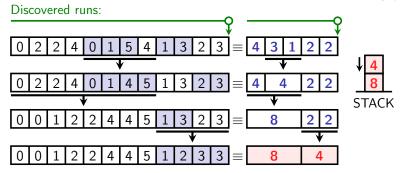
- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack

  - merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs

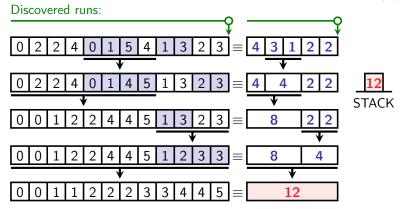


- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack

  - merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack
  - 2 merge the top 1st and 2nd runs
  - 3 merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



- Maintain a stack of runs
- Until the array is sorted, either:
  - discover & push a new run onto the stack
  - 2 merge the top 1st and 2nd runs
  - merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs

#### Key ideas:

• Each run r pays its share of the total merge cost

#### Key ideas:

- Each run r pays
  - $\triangleright$   $\mathcal{O}(r)$  to enter the stack (entry phase)
  - ▶ r to increase its **bit length (growth phase)** bit length of r:  $\ell = \lfloor \log_2(r) \rfloor$

- Each run r pays
  - $\mathcal{O}(r)$  during its own run entry phase
  - at most  $r\lceil \log_2(n/r) \rceil$  during the growth phases
- Total merge cost of  $n\mathcal{H} + \mathcal{O}(n)$

#### Key ideas:

- Each run r pays
  - $ightharpoonup \mathcal{O}(r)$  to **enter** the stack (**entry phase**)
  - ▶ r to increase its **bit length (growth phase)** bit length of r:  $\ell = \lfloor \log_2(r) \rfloor$
- Entry phase:

#### Cost analysis:

- Each run r pays
  - $\mathcal{O}(r)$  during its own **run entry phase**
  - at most  $r\lceil \log_2(n/r) \rceil$  during the growth phases
- Total merge cost of  $n\mathcal{H} + \mathcal{O}(n)$

Pushed run r  $r_k$   $\vdots$   $r_{h-1}$   $r_{h-2}$ 

STACK

ľЗ

#### Key ideas:

- Each run r pays
  - $\mathcal{O}(r)$  to enter the stack (entry phase)
  - ▶ r to increase its **bit length** (**growth phase**) bit length of r:  $\ell = \lfloor \log_2(r) \rfloor$
- Entry phase:

# Rule 🚳 Pushed run $r_k$ Run entry collapse ľЗ

STACK

- Each run r pays
  - $\mathcal{O}(r)$  during its own run entry phase
  - at most  $r\lceil \log_2(n/r)\rceil$  during the growth phases
- Total merge cost of  $n\mathcal{H} + \mathcal{O}(n)$

#### Key ideas:

- Each run r pays
  - $\mathcal{O}(r)$  to enter the stack (entry phase)
  - ▶ r to increase its **bit length (growth phase)** bit length of r:  $\ell = \lfloor \log_2(r) \rfloor$
- Entry phase:

Pushed run  $r_h$ Merged run  $r_{h-1}$   $r_{h-2}$   $\vdots$ 

- Each run r pays
  - $\mathcal{O}(r)$  during its own **run entry phase**
  - at most  $r\lceil \log_2(n/r) \rceil$  during the growth phases
- Total merge cost of  $n\mathcal{H} + \mathcal{O}(n)$

#### Key ideas:

- Each run r pays
  - $\mathcal{O}(r)$  to enter the stack (entry phase)
  - ▶ r to increase its **bit length** (**growth phase**) bit length of r:  $\ell = \lfloor \log_2(r) \rfloor$
- Entry phase: ensure that
  - r pays for every merge
  - ▶  $(r_i)_{i\geqslant 1}$  has **exponential** decay when r is pushed
  - ▶ runs **smaller** than *r* are merged

# Run entry collapse ľЗ

 $r_k$ 

STACK

Pushed run

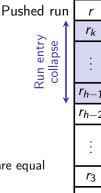
- Each run r pays
  - $\mathcal{O}(r)$  during its own run entry phase
  - at most  $r\lceil \log_2(n/r) \rceil$  during the growth phases
- Total merge cost of  $n\mathcal{H} + \mathcal{O}(n)$

#### Key ideas:

- Each run r pays
  - $\mathcal{O}(r)$  to enter the stack (entry phase)
  - ▶ r to increase its **bit length** (**growth phase**) bit length of r:  $\ell = \lfloor \log_2(r) \rfloor$
- Entry phase: ensure that
  - r pays for every merge
  - $(r_i)_{i\geqslant 1}$  has **exponential** decay when r is pushed
  - runs smaller than r are merged
- Growth phase: ensure that
  - $ightharpoonup r_i$  and  $r_{i+1}$  are merged only if their **bit lengths** are equal

#### Cost analysis:

- Each run r pays
  - $\mathcal{O}(r)$  during its own run entry phase
  - at most  $r\lceil \log_2(n/r) \rceil$  during the growth phases
- Total merge cost of  $n\mathcal{H} + \mathcal{O}(n)$



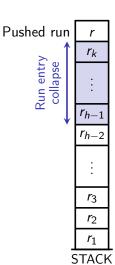
 $r_2$ 

STACK

#### Key ideas:

- Each run r pays
  - $\mathcal{O}(r)$  to enter the stack (entry phase)
  - ▶ r to increase its **bit length (growth phase)** bit length of r:  $\ell = \lfloor \log_2(r) \rfloor$
- Entry phase: ensure that
  - r pays for every merge
  - $(\ell_i)_{i\geqslant 1}$  is **decreasing** when r is pushed
  - ▶ runs  $r_i$  with  $\ell_i \leq \ell$  are merged
- Growth phase: ensure that
  - $r_i$  and  $r_{i+1}$  are merged only if  $\ell_i = \ell_{i+1}$

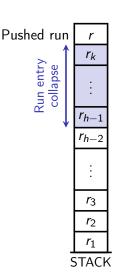
- Each run r pays
  - $\mathcal{O}(r)$  during its own run entry phase
  - at most  $r\lceil \log_2(n/r) \rceil$  during the growth phases
- Total merge cost of  $n\mathcal{H} + \mathcal{O}(n)$



#### Key ideas:

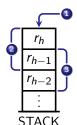
- Each run r pays
  - $\mathcal{O}(r)$  to enter the stack (entry phase)
  - ▶ r to increase its **bit length (growth phase)** bit length of r:  $\ell = \lfloor \log_2(r) \rfloor$
- Entry phase: ensure that
  - r pays for every merge
  - $(\ell_i)_{i\geqslant 1}$  is **decreasing** when r is pushed  $\triangleright$
  - ▶ runs  $r_i$  with  $\ell_i \leq \ell$  are merged
- Growth phase: ensure that
  - $r_i$  and  $r_{i+1}$  are merged only if  $\ell_i = \ell_{i+1}$   $\rho \rho$

- Each run r pays
  - $\mathcal{O}(r)$  during its own run entry phase
  - at most  $r\lceil \log_2(n/r) \rceil$  during the growth phases
- Total merge cost of  $n\mathcal{H} + \mathcal{O}(n)$



#### Choice rules for options

- discover & push a new run length onto the stack
- 2 merge the top 1<sup>st</sup> and 2<sup>nd</sup> runs
- merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



#### Choice algorithm

```
if \ell_h \geqslant \ell_{h-2} or \ell_{h-1} \geqslant \ell_{h-2}: choose ③
```

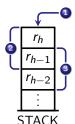
else if  $\ell_h \geqslant \ell_{h-1}$ : choose ②

else: choose ① (or ② if ① is unavailable)

where 
$$\ell_i = \lfloor \log_2(r_i) \rfloor$$

#### Choice rules for options

- discover & push a new run length onto the stack
- 2 merge the top 1<sup>st</sup> and 2<sup>nd</sup> runs
- merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



#### Choice algorithm

```
if \ell_h \geqslant \ell_{h-2} or \ell_{h-1} \geqslant \ell_{h-2}: choose 3
```

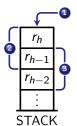
else if  $\ell_h \geqslant \ell_{h-1}$ : choose ②

else: choose ① (or ② if ① is unavailable)

where 
$$\ell_i = \lfloor \log_2(r_i) \rfloor$$

#### Choice rules for options

- discover & push a new run length onto the stack
- merge the top 1<sup>st</sup> and 2<sup>nd</sup> runs
- merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



#### Choice algorithm

```
if \ell_h \geqslant \ell_{h-2} or \ell_{h-1} \geqslant \ell_{h-2}: choose ③
```

else if  $\ell_h \geqslant \ell_{h-1}$ : choose ②

else: choose ① (or ② if ① is unavailable)

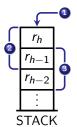
where 
$$\ell_i = \lfloor \log_2(r_i) \rfloor$$

#### Bit-length constraints:

- $\ell_1 > \ell_2 > \ldots > \ell_{h-2} \geqslant \ell_{h-1}$  (induction)
- $\ell_1 > \ell_2 > \ldots > \ell_h$  on run push ho
- $\ell_{h-1} \geqslant \ell_h$  and  $\ell_{h-2} > \ell_h$  during growth (induction)  $\rho \rho$

#### Choice rules for options

- discover & push a new run length onto the stack
- merge the top 1<sup>st</sup> and 2<sup>nd</sup> runs
- merge the top 2<sup>nd</sup> and 3<sup>rd</sup> runs



#### Choice algorithm

if  $\ell_h \geqslant \ell_{h-2}$  or  $\ell_{h-1} \geqslant \ell_{h-2}$ : choose ③

else if  $\ell_h \geqslant \ell_{h-1}$ : choose ②

else: choose ① (or ② if ① is unavailable)

where 
$$\ell_i = \lfloor \log_2(r_i) \rfloor$$

#### Bit-length constraints:

- $\ell_1 > \ell_2 > \ldots > \ell_{h-2} \geqslant \ell_{h-1}$  (induction)
- $\ell_1 > \ell_2 > \ldots > \ell_h$  on run push ho
- $\ell_{h-1} \geqslant \ell_h$  and  $\ell_{h-2} > \ell_h$  during growth (induction)  $\rho \rho$



#### Conclusion

- TimSort is good in practice and in theory:  $\mathcal{O}(n + n\mathcal{H})$  merge cost
- Adaptive ShiversSort is better than and very similar to TimSort

#### Conclusion

- TimSort is good in practice and in theory:  $\mathcal{O}(n+n\mathcal{H})$  merge cost
- Adaptive ShiversSort is better than and very similar to TimSort

#### Some references:

[1]	Optimal computer search trees and variable-length alphabetical codes,	
	Hu & Tucker	(1971)
[2]	A new algorithm for minimum cost binary trees, Garsia & Wachs	(1973)
[3]	Tim Peters' description of TimSort,	
	<pre>svn.python.org/projects/python/trunk/Objects/listsort.txt</pre>	(2001)
[4]	On compressing permutations and adaptive sorting, Barbay & Navarro	(2013)
[5]	OpenJDK's java.utils.Collection.sort() is broken, de Gouw et al.	(2015)
[6]	Merge strategies: from merge sort to TimSort, Auger et al.	(2015)
[7]	On the worst-case complexity of TimSort, Auger et al.	(2018)
[8]	Nearly-optimal mergesorts, Munro & Wild	(2018)
[9]	Strategies for stable merge sorting, Buss & Knop	(2019)
[10]	Adaptive ShiversSort: an alternative sorting algorithm, Jugé	(2020)

