

Compsci 201
P4 Analysis
Annie Lin and Eden Deng

(Please see runtime results for DNABenchmark and AnalysisDNA at the bottom)

Question 1: are the benchmark timings for StringStrand consistent with the explanation below that the time to execute cutAndSplice is $O(b^2S)$?

Note that the value of **b** is half the number of calls to append since each cut (except the first) is modeled by two calls of append in the method **cutAndSplice** -- see the code. This means that b^2 will be constant in the benchmark, but S will vary.

Answer:

The benchmark timings for StringStrand are consistent with $O(b^2S)$ cutAndSplice runtime efficiency.

b is the occurrences of the enzyme, and S is the length of the splicee. In each run, b remains constant, and as S doubles, runtime also nearly doubles. The trend is increasingly obvious as S becomes larger. For example, when S increases from 131,072 to 262,144, time increases from 11.154 to 23.884; when S increases from 262,144 to 524,288, time increases from 23.884 to 47.238.

As seen in the results from AnalysisDNA, where the number of occurrences of restriction enzyme sites is increasing linearly with each trial holding splicee size constant, the runtime for StringStrand is consistent with $O(b^2S)$. When b doubles from the first to second trial, the runtime increases by close to a factor of four (4.810 to 11.206). Similarly, when b doubles again from the second trial to the fourth trial, the runtime increases by approximately a factor of four (11.206 to 42.432).

Question 2: are the benchmark timings for **StringBuilderStrand** consistent with the explanation below that the time to execute cutAndSplice is **O(bS)**?

Note that the value of **b** is half the number of calls to append since each cut (except the first) is modeled by two calls of append in the method **cutAndSplice** -- see the code.

Answer:

The benchmark timings for StringBuilderStrand are consistent with $O(bS)$ cutAndSplice runtime efficiency.

In each run, **b** remains constant, and as **S** doubles, runtime also nearly doubles. For example, when **S** increases from 131,072 to 262,144, time increases from 0.252 to 0.417. (The trend only becomes obvious when **S** becomes larger. When **S** is small, time is always less than 0.05 seconds, so the inconsistency is negligible.

In AnalysisDNA where **b** is increasing linearly with each trial, the results are consistent with $O(bS)$ runtime complexity. For instance, when **b** doubles from the fourth trial to the eighth trial, the runtime also roughly doubles from 0.071 to 0.191, as expected.

Question 3: Explain why the time for LinkStrand does not change much at all over all the runs in the benchmark program. Explain why you think memory is exhausted at the specific strand size you see in your timings -- as compared to exhaustion for String and StringBuilder.

Answer:

In LinkStrand, each time when cutAndSplice is implemented, no new string is generated; instead, the nodes undergoing change point to the same splicee using pointers (which is constant efficiency).

Memory is exhausted at strand size 536,870,912, which is 1,000 times larger than that of StringStrand 524,288, and 2,000 times larger than that of StringBuilderStrand 262,144. LinkStrand implementations use a lot less memory because pointers are used in cutAndSplice and no new string is generated; on the other hand, StringStrand and StringBuilderStrand do generate new strings, which causes memory to be exhausted a lot earlier.

Running DNA Benchmark

StringStrand

dna length = 4,639,221

cutting at enzyme gaattc

| Class | splicee | recomb | time | appends |
|---------------|---------|-------------|--------|---------|
| ----- | | | | |
| StringStrand: | 256 | 4,800,471 | 0.446 | 1290 |
| StringStrand: | 512 | 4,965,591 | 0.457 | 1290 |
| StringStrand: | 1,024 | 5,295,831 | 0.523 | 1290 |
| StringStrand: | 2,048 | 5,956,311 | 0.544 | 1290 |
| StringStrand: | 4,096 | 7,277,271 | 0.671 | 1290 |
| StringStrand: | 8,192 | 9,919,191 | 0.924 | 1290 |
| StringStrand: | 16,384 | 15,203,031 | 1.462 | 1290 |
| StringStrand: | 32,768 | 25,770,711 | 2.580 | 1290 |
| StringStrand: | 65,536 | 46,906,071 | 5.924 | 1290 |
| StringStrand: | 131,072 | 89,176,791 | 11.154 | 1290 |
| StringStrand: | 262,144 | 173,718,231 | 23.884 | 1290 |
| StringStrand: | 524,288 | 342,801,111 | 47.238 | 1290 |

StringBuilderStrand

dna length = 4,639,221

cutting at enzyme gaattc

| Class | splicee | recomb | time | appends |
|----------------------|---------|-------------|-------|---------|
| ----- | | | | |
| StringBuilderStrand: | 256 | 4,800,471 | 0.041 | 1290 |
| StringBuilderStrand: | 512 | 4,965,591 | 0.031 | 1290 |
| StringBuilderStrand: | 1,024 | 5,295,831 | 0.027 | 1290 |
| StringBuilderStrand: | 2,048 | 5,956,311 | 0.030 | 1290 |
| StringBuilderStrand: | 4,096 | 7,277,271 | 0.024 | 1290 |
| StringBuilderStrand: | 8,192 | 9,919,191 | 0.007 | 1290 |
| StringBuilderStrand: | 16,384 | 15,203,031 | 0.012 | 1290 |
| StringBuilderStrand: | 32,768 | 25,770,711 | 0.036 | 1290 |
| StringBuilderStrand: | 65,536 | 46,906,071 | 0.033 | 1290 |
| StringBuilderStrand: | 131,072 | 89,176,791 | 0.252 | 1290 |
| StringBuilderStrand: | 262,144 | 173,718,231 | 0.417 | 1290 |

LinkStrand

dna length = 4,639,221

cutting at enzyme gaattc

| Class | splicee | recomb | time | appends |
|-------------|-------------|-----------------|-------|---------|
| LinkStrand: | 256 | 4,800,471 | 0.027 | 1290 |
| LinkStrand: | 512 | 4,965,591 | 0.021 | 1290 |
| LinkStrand: | 1,024 | 5,295,831 | 0.004 | 1290 |
| LinkStrand: | 2,048 | 5,956,311 | 0.006 | 1290 |
| LinkStrand: | 4,096 | 7,277,271 | 0.006 | 1290 |
| LinkStrand: | 8,192 | 9,919,191 | 0.004 | 1290 |
| LinkStrand: | 16,384 | 15,203,031 | 0.004 | 1290 |
| LinkStrand: | 32,768 | 25,770,711 | 0.003 | 1290 |
| LinkStrand: | 65,536 | 46,906,071 | 0.004 | 1290 |
| LinkStrand: | 131,072 | 89,176,791 | 0.004 | 1290 |
| LinkStrand: | 262,144 | 173,718,231 | 0.004 | 1290 |
| LinkStrand: | 524,288 | 342,801,111 | 0.010 | 1290 |
| LinkStrand: | 1,048,576 | 680,966,871 | 0.004 | 1290 |
| LinkStrand: | 2,097,152 | 1,357,298,391 | 0.004 | 1290 |
| LinkStrand: | 4,194,304 | 2,709,961,431 | 0.006 | 1290 |
| LinkStrand: | 8,388,608 | 5,415,287,511 | 0.007 | 1290 |
| LinkStrand: | 16,777,216 | 10,825,939,671 | 0.005 | 1290 |
| LinkStrand: | 33,554,432 | 21,647,243,991 | 0.007 | 1290 |
| LinkStrand: | 67,108,864 | 43,289,852,631 | 0.005 | 1290 |
| LinkStrand: | 134,217,728 | 86,575,069,911 | 0.009 | 1290 |
| LinkStrand: | 268,435,456 | 173,145,504,471 | 0.006 | 1290 |
| LinkStrand: | 536,870,912 | 346,286,373,591 | 0.083 | 1290 |

Running AnalysisDNA

StringStrand

dna length = 4,639,221
cutting at enzyme gaattc

| Class | splicee | recomb | time | appends |
|---------------|---------|------------|---------|---------|
| ----- | | | | |
| StringStrand: | 4,096 | 14,554,542 | 4.810 | 2580 |
| StringStrand: | 4,096 | 21,831,813 | 11.206 | 3870 |
| StringStrand: | 4,096 | 29,109,084 | 24.138 | 5160 |
| StringStrand: | 4,096 | 36,386,355 | 42.432 | 6450 |
| StringStrand: | 4,096 | 43,663,626 | 57.036 | 7740 |
| StringStrand: | 4,096 | 50,940,897 | 68.456 | 9030 |
| StringStrand: | 4,096 | 58,218,168 | 91.178 | 10320 |
| StringStrand: | 4,096 | 65,495,439 | 120.574 | 11610 |
| StringStrand: | 4,096 | 72,772,710 | 159.558 | 12900 |
| StringStrand: | 4,096 | 80,049,981 | 207.129 | 14190 |

StringBuilderStrand

dna length = 4,639,221
cutting at enzyme gaattc

| Class | splicee | recomb | time | appends |
|----------------------|---------|------------|-------|---------|
| ----- | | | | |
| StringBuilderStrand: | 4,096 | 7,277,271 | 0.079 | 1290 |
| StringBuilderStrand: | 4,096 | 14,554,542 | 0.124 | 2580 |
| StringBuilderStrand: | 4,096 | 21,831,813 | 0.198 | 3870 |
| StringBuilderStrand: | 4,096 | 29,109,084 | 0.071 | 5160 |
| StringBuilderStrand: | 4,096 | 36,386,355 | 0.140 | 6450 |
| StringBuilderStrand: | 4,096 | 43,663,626 | 0.246 | 7740 |
| StringBuilderStrand: | 4,096 | 50,940,897 | 0.212 | 9030 |
| StringBuilderStrand: | 4,096 | 58,218,168 | 0.191 | 10320 |
| StringBuilderStrand: | 4,096 | 65,495,439 | 0.102 | 11610 |
| StringBuilderStrand: | 4,096 | 72,772,710 | 0.457 | 12900 |

LinkStrand

dna length = 4,639,221
cutting at enzyme gaattc

| Class | splicee | recomb | time | appends |
|-------|---------|--------|------|---------|
|-------|---------|--------|------|---------|

| | | | | |
|-------------|-------|------------|-------|-------|
| LinkStrand: | 4,096 | 7,277,271 | 0.070 | 1290 |
| LinkStrand: | 4,096 | 14,554,542 | 0.069 | 2580 |
| LinkStrand: | 4,096 | 21,831,813 | 0.017 | 3870 |
| LinkStrand: | 4,096 | 29,109,084 | 0.048 | 5160 |
| LinkStrand: | 4,096 | 36,386,355 | 0.034 | 6450 |
| LinkStrand: | 4,096 | 43,663,626 | 0.049 | 7740 |
| LinkStrand: | 4,096 | 50,940,897 | 0.039 | 9030 |
| LinkStrand: | 4,096 | 58,218,168 | 0.079 | 10320 |
| LinkStrand: | 4,096 | 65,495,439 | 0.143 | 11610 |
| LinkStrand: | 4,096 | 72,772,710 | 0.073 | 12900 |