# CS4040 Review Form

**Student:** Sheikh Usman Ali
**Title:** Benchmarking Computational speed of 1-D FFT algorithms

**Marker:** Konrad Dryja

December 13, 2019
University of Aberdeen

# 1 INTRODUCTION, BACKGROUND, RESEARCH QUESTION

## 1.1 Does the Introduction present the context and "big picture"?

Student introduced notion of Fast Fourier Transform and explained where it is used, but never actually explained how and what problems the algorithms are solving. Perhaps it'd be useful to include a sentence or two on the nature of problems that FFT are used in. Nevertheless, the entire paper has been correctly condensed and summarised in the introduction and I agree that the algorithm itself is not the subject of the paper, but rather its performance.

## 1.2 Does related work contain adequate references to related work and discussion of that work?

Yes, plenty of examples from industry attempting to benchmark the algorithm.

## 1.3 Are references properly formatted and complete?

Yes, although I feel like some extra references would be needed too, student sometimes makes claims without referencing them to any other source (e.g. "There are other Benchmarking open source and private results[...]", citing or at least linking one of them for reference would be better).

Also Figure 2 was never referenced. If it has been produced for the purposes of this project, it should contain more explanation, otherwise it should be cited.

## 1.4 Is a research question presented and tied into the literature review/background?

I couldn't find a clearly stated research question in the third section - or in any of the earlier ones. Reading through "Research Question" section (sec 3), it feels more like experimental design, rather than trying to identify the claim that we are trying to prove or disprove. Student describes *how* to measure, but never actually mentions what we are trying to find out. One could deduce from the content presented that *general* performance is measured, although I don't think it's something that reader should be deducing themselves - rather it should be clearly stated by the author.

## 1.5 COMMENTS AND MARK

- The two biggest things that were missing and / or skipped in my opinion is slightly too advanced introduction - which might be hard to understand without any background or knowledge of FFT algorithms and the fact that Research Question section feels more like description of the experiment, forcing the reader to assume things about the purpose of the paper.

- There were also couple of very minor stylistic issues (e.g. quote in section 2, with mixed italics in it?) and mentioned above suggestions regarding references. Other than that, very advanced content.

**CGS B:** Good, with some problems.

# 2 EXPERIMENTAL DESIGN

## 2.1 Are the hypotheses appropriate to the research question?

Considering my comment in the previous section, it's hard to say, since the Research Question was never stated. Although judging by the description in section 3, the hypotheses are appropriate.

## 2.2 Are the systems or algorithms being evaluated properly explained?

When describing the operating systems, student said that Pentium 4 is running Linux 2.4.25 and i7 CPU is running Ubuntu 18.04 - but one thing is a kernel and then other is a distribution, which can't really be compared, since Ubuntu 18.04 - in theory - could also be running kernel 2.4.25.

It could also be useful to list the RAM and Storage capabilities of each machine, since I/O speeds (of both volatile and non-volatile memories) could affect the runtime of algorithms.

Also the version of BenchFFT was not mentioned, which makes it hard to reproduce the test should BenchFFT be updated to a different version, thus in theory changing its features and potential test outcomes.

## 2.3 Is the experiment described in sufficient detail to be replicated?

Yes, although sometimes vague ("approximately 2-3 days")

## 2.4 Does the experimental design enable the hypotheses to be adequately tested?

Yes, BenchFFT should be able to measure raw performance of both machines and see which one is performing better by comparing explained metrics (Mflops).

## 2.5 Are other, potentially influencing, factors controlled for/managed?

Some, yes. Student mentions running the benchmarking suite with the same parameters and without any additional compiler flags. It would also be nice to reduce potential background noise, i.e. other processes that might be taking CPU time and thus spoiling the results.

Other than that, it's nice to average over several runs.

## 2.6 Are representative test cases selected? If human subjects are used, are they appropriate?

Yes, I like the normalisation to obtain comparable values.

## 2.7 COMMENTS AND MARK

- "According to the FFTW team" - could use a reference.

- Also some very weird, randomly uppercased words in the middle of sentences, e.g.: "Correlation", "Hardware", "Design". In my opinion, it unnecessary stands out and doesn't make sense grammatically.

- At its current state, it might not be possible to clearly reproduce the test, as some information is missing, such as more detailed hardware information or versions of benchmarking software. Although the experiment was thoroughly explained, which makes sense and should yield data allowing to tackle our hypotheses.

**CGS B:** Quite good, although shame about reproducibility.

# 3 RESULTS

## 3.1 Are there an appropriate number of figures presenting results of the experiments?

Yes, lots of graphs.

Student mentions that "tables are excluded from the report". I understand the sentiment, although maybe it would still be nice to include them in the appendix.

## 3.2 Are appropriate statistical tests performed, and p values presented?

Student mentions lot of p-values, although only once states the actual value, normally deciding to only state that it's smaller than 0.05 - without precise numbers. Perhaps to increase the credibility of the tests, it'd be nice to list the exact values.

It's also never explained *what* the p-test is performed against, only that they have been performed.

## 3.3 Is a qualitative analysis given of mistakes and poor performance?

No, student once encountered p-value of 0.8028, but never explained why this might have been the case and how could this been avoided.

## 3.4 COMMENTS AND MARK

- I'd focus more on the case where the results weren't statistically significant and write an extra paragraph as to what could be the reason.

- Also, small thing, student mentions $R^2$ value, but never actually explains what it exactly means.

- Other than that, lot of complex graphs with appropriate explanations.

**CGS A/B:** Good

# 4 DISCUSSION AND CONCLUSION

## 4.1 Does the discussion follow from the results?

Yes, it's a direct continuation of the previous section

## 4.2 Does the discussion present key insights from the result?

Yes, although it feels like the conclusion focuses more on the comparison of various algorithms, as opposed to comparing the same algorithms on different machines (as per stated hypotheses in section 4).

## 4.3 Is sensible future work proposed?

Yes, student suggests looking into memory access times or considering different implementations of algorithms to further reduce the noise.

## 4.4 Does the conclusion summarise the key findings of the report?

Kind of, starts fine, mentioning that all algorithms perform well, but then sidetracks to the fact that benchmarking is required and can be difficult for an end-user.

## 4.5 COMMENTS AND MARK

- The "Discussion" section seems to be slightly incoherent, often jumping between different topics (e.g., one sentence talks about multi-core vs single-core performance and immediately after starts talking about GPUs, the flow doesn't feel natural).

- I also disagree with the last sentence in section 6, as reproducibility mostly concerns *the exact* attempt of recreating the test, so using the identical hardware is implied.

**CGS B:** Good.

# 5 OVERALL COMMENTS AND MARK

- The report tackled a quite heavy topic of benchmarking complex algorithms. It draws appropriate conclusions, although sometimes gets lost and sidetracks to different issues, which weren't defined as in-scope for the report. I think it was mostly caused by the fact that research question wasn't ever strictly stated, thus mentioning anything became a possibility.

- The paper also has some spelling and grammar mistakes, often capitalising random words in sentences which hurts readability. There is also some inconsistencies with spelling of brand names (e.g. sometimes spelling "i7", sometimes "i-7" and sometimes "I7")

- Finally, the report draws conclusions, but never relates them to the stated hypotheses and instead discusses efficiency of various algorithms - which isn't in-scope, from what I understood.

I still think that it was a good report, where the student choose a quite heavy and hectic subject to research, thus limiting the possible maneuvers, so aforementioned problems naturally sneaked in. Results were appropriate and satisfactory.

**Overall mark: B2**