APPROXIMATE ORACLES AND SYNERGY IN SOFTWARE ENERGY SEARCH SPACES

Bobby R. Bruce, Justyna Petke, Mark Harman, and Earl T. Barr
Reducing Energy Consumption Using Genetic Improvement

Bobby R. Bruce
University College London
London
United Kingdom
r.bruce@cs.ucl.ac.uk

Justyna Petke
University College London
London
United Kingdom
j.petke@ucl.ac.uk

Mark Harman
University College London
London
United Kingdom
mark.harman@ucl.ac.uk

ABSTRACT
Genetic Improvement (GI) is an area of Search Based Software Engineering which seeks to improve software's non-functional properties by treating program code as if it were genetic material which is then evolved to produce more optimal solutions. Historically, the majority of focus has been on optimizing program's execution time which, though important, is only one of many non-functional targets. The growth in mobile computing, cloud computing infrastructure, and ecological concerns are forcing developers to focus on the energy their software consumes. We report on investigations into using GI to automatically find more energy efficient versions of the MiniSAT Boolean satisfiability solver when specialising for these downstream applications. Our results find that GI can successfully be used to reduce energy consumption by up to 25%.

Categories and Subject Descriptors
D.2 [Software]: Software Engineering

Keywords
Search based software engineering, SBSE, genetic improvement, GI, optimisation, energy optimisation, energy efficiency, energy consumption, Boolean satisfiability

1. INTRODUCTION
Less than a decade ago the quality of software (outside of end-user design preferences) could broadly be described as the extent to which software met its specification while minimising the prevalence of bugs and usage of traditional computer resources such as CPU time and memory allocation. The growth in two new technologies, mobile computing devices and cloud services, has led to a new environment for software engineers where they must now consider the energy an application consumes. The quality of software is now measured in Joules, as well as bug counts, seconds, and megabytes. As presented at here are more smartphones in the world than personal computers [22], each containing a limited store of energy between changes that must be used efficiently. The energy required to run large server clusters has grown considerably in the last decade, estimated to be between 1.1% to 1.3% of global electricity consumption in 2010 [26], putting strain on energy suppliers and the budgets of those responsible for purchasing this energy [7]. The total ICT infrastructure generated 1.9% of global CO2 emissions in 2011 [5] (larger than the entire United Kingdom estimated at 1.47% for the 2010-2014 period [42]) indicating that computer science has a role to play in mitigating climate change.

Thus we believe it important that software engineers find ways of programming computers with energy efficiency in mind to appease the demands from consumers for longer battery life, from companies to reduce their energy bills, and from society’s desire to minimise humanity’s impact on the environment.

One of the largest hurdles in producing energy-efficient software is in the developer’s disconnect between the source code they write and the energy that will be consumed from the compiled product they deliver [35]. Without a deep understanding of how a particular compiler works, along with an equally deep understanding of how much energy a given instruction will consume, the problem remains difficult for many developers. It has been found that metrics previously believed to guide developers to more energy efficient solutions are, in reality, poor at doing so [36]. Subtle changes, such as introducing inline methods [41], swapping API implementations [35], and constructing semantically equivalent (but structurally inequivalent) algorithms [6] have all been shown to influence energy consumption. However this influence is difficult to determine outside of the ad hoc and inefficient process of trial-and-error. Tools have been developed to guide users to energy-efficient areas of their software [7, 11, 30, 36] though the developer retains responsibility for rectifying these inefficiencies.

We suggest that the most under-explored method of decreasing software’s energy consumption lies in automated processes. Such processes would allow developers to focus solely on meeting the specification requirements with worries about non-functional attributes like energy consumption left to an algorithm capable of refactoring software to a more optimal state.

Genetic Improvement (GI) [20, 25, 27, 28, 29, 36, 37, 45, 44] is a Search Based Software Engineering (SBSE) technique [21] which treats program code as if it were genetic material that can then be evolved to produce optimised solutions. GI has previously been found effective at optimising

Some “Threats to Validity”

➤ Only for 1 (very small) piece of software
➤ Limited number of runs
➤ Evidence our current approach to Genetic Improvement was not ideal
➤ Indirect energy measurement (Intel Power Gadget)

Published at GECCO 2015

Used Genetic Improvement to reduce the energy consumption of MiniSAT

Bobby R. Bruce
FOLLOWUP: LET'S DO THE SAME, BUT BIGGER!

Bobby R. Bruce
FOLLOWUP: LET'S DO THE SAME, BUT BIGGER!

Real, direct energy measurements!

Bobby R. Bruce
FOLLOWUP: LET’S DO THE SAME, BUT BIGGER!

Real, direct energy measurements!

More applications!

Bobby R. Bruce
FOLLOWUP: LET’S DO THE SAME, BUT BIGGER!

Real, direct energy measurements!

More applications!

Larger applications!

Bobby R. Bruce
FOLLOWUP: LET’S DO THE SAME, BUT BIGGER!

Real, direct energy measurements!

More applications!

Larger applications!

More evaluations!

Bobby R. Bruce
THE FRAMEWORK WE DEVELOPED

Software + Input

Master

Node

Measurement Board

Controller

Node

Output + Energy

Energy Measurement Board

Bobby R. Bruce
THE COPY, DELETE, AND REPLACE OPERATORS

Bobby R. Bruce
THE COPY, DELETE, AND REPLACE OPERATORS

<IF_fileA_40>

Bobby R. Bruce
THE COPY, DELETE, AND REPLACE OPERATORS

<IF_fileA_40>

<for1_fileB_6><for1_fileB_78>

Bobby R. Bruce
THE COPY, DELETE, AND REPLACE OPERATORS

<IF_fileA_40>

<for1_fileB_6><for1_fileB_78>

<fileA_101>+<fileA_15>

Bobby R. Bruce
THE COPY, DELETE, AND REPLACE OPERATORS

<for1_fileB_6> <for1_fileB_78> <IF_fileA_40>

<fileA_101> + <fileA_15>

Bobby R. Bruce
THE COPY, DELETE, AND REPLACE OPERATORS

<for1_fileB_6> <for1_fileB_78> <IF_fileA_40> <fileA_101> + <fileA_15>

Bobby R. Bruce
THE GREAT GENETIC IMPROVEMENT CYCLE

Source Code

Representation

Population

Better Source Code

Altered Representation

Evaluation

Selection

Mutation

Crossover

Bobby R. Bruce
APPLICATIONS STUDIED

7ZIP
Bodytrack
Ferret
PARSEC
OMXplayer

Bobby R. Bruce
## INITIAL INVESTIGATION

<table>
<thead>
<tr>
<th>App</th>
<th>Pop Size</th>
<th>Num Gen</th>
<th>Improvement?</th>
</tr>
</thead>
<tbody>
<tr>
<td>7zip</td>
<td>100</td>
<td>20</td>
<td>Nope</td>
</tr>
<tr>
<td>Bodytrack</td>
<td>100</td>
<td>20</td>
<td>Nope</td>
</tr>
<tr>
<td>Bodytrack</td>
<td>100</td>
<td>50</td>
<td>Nope</td>
</tr>
</tbody>
</table>

_Bobby R. Bruce_
Maybe 7zip and Bodytrack can’t be optimised very much?
Maybe 7zip and Bodytrack can’t be optimised very much?

Maybe our search algorithm isn’t suitable?
Maybe 7zip and Bodytrack can’t be optimised very much?

Maybe our search algorithm isn’t suitable?

How effective can an individual modification be?
Maybe 7zip and Bodytrack can’t be optimised very much?

Maybe our search algorithm isn’t suitable?

How effective can an individual modification be?

If an individual modification’s effectiveness is small, can it be detected?

Bobby R. Bruce
What does the search space look like?
RESEARCH QUESTIONS

RQ1: Measurement: What variance occurs when measuring energy consumption?

RQ1a: What is the variance when measuring using a single energy measurement device?

RQ1b: What is the variance in direct energy measurements across multiple devices?

RQ1c: What is the variance in proportional energy changes across multiple devices?

RQ2: Improvement: What additional energy improvement can be achieved when using approximate oracles in place of exact oracles?

RQ3: Synergy: How frequently do synergistic or antagonistic effects occur when combining known effective modifications?
# The Software Targets

| App       | Modifiable\n| Lines of code | What does “passed” mean? | What can be approximated? |
|-----------|----------------|--------------------------|---------------------------|
| 7zip      | 2,524          | Creates a valid .7z file | Compression Rate          |
| Bodytrack | 1,030          | Creates a non-empty, readable output | The accuracy of the tracking |
| Ferret    | 5,032          | Creates a non-null, readable list of images | The ranking |
| OMXPlayer | 5,184          | Outputs data to the HDMI port | The video quality          |
RQ1, MEASUREMENT

WHAT VARIANCE OCCURS WHEN MEASURING ENERGY CONSUMPTION?

RQ1a: What is the variance when measuring using a single measurement device?

Bobby R. Bruce
RQ1, MEASUREMENT

WHAT VARIANCE OCCURS WHEN MEASURING ENERGY CONSUMPTION?

*RQ1b: What is the variance in direct energy measurements across multiple devices?*
RQ1, MEASUREMENT

WHAT VARIANCE OCCURS WHEN MEASURING ENERGY CONSUMPTION?

RQ1c: What is the variance in proportional energy changes across multiple devices?

Bobby R. Bruce
RQ1 CONCLUSIONS

- The MAGEEC Energy Measurement boards have provided us with a framework capable of acceptable precision but low accuracy.

- While the readings, when reported in Joules, vary wildly, the proportional changes are relatively stable. Thus proportional change (i.e. a percentage increase or decrease) is recorded.

- The variation of energy readings within devices appear acceptably small to detect meaningful changes when they occur.
RQ2, IMPROVEMENT

WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?
RQ2, IMPROVEMENT

WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?
RQ2, IMPROVEMENT

WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?

Software

Candidate Modification list

Unmodified energy data

Measure Energy x30

For each test

Test Set

Bobby R. Bruce
RQ2, IMPROVEMENT

WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?
RQ2B, IMPROVEMENT

WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?

---

**Bodytrack**

*Original*  

*33% Energy Reduction*
RQ2B, IMPROVEMENT

WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?

<table>
<thead>
<tr>
<th>Energy reduction</th>
<th>Approximation Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.69%</td>
<td>0.000</td>
</tr>
<tr>
<td>19.26%</td>
<td>0.131</td>
</tr>
<tr>
<td>27.97%</td>
<td>0.170</td>
</tr>
<tr>
<td>29.13%</td>
<td>0.192</td>
</tr>
<tr>
<td>33.69%</td>
<td>0.452</td>
</tr>
</tbody>
</table>

*Bodytrack*

*Bobby R. Bruce*
What additional energy improvement can be achieved when using approximate test oracles in place of exact test oracles?

### 7zip

<table>
<thead>
<tr>
<th>Energy reduction</th>
<th>Approximation Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.08%</td>
<td>$3.93 \times 10^{-4}$</td>
</tr>
<tr>
<td>12.29%</td>
<td>0.072</td>
</tr>
<tr>
<td>13.17%</td>
<td>0.102</td>
</tr>
<tr>
<td>48.30%</td>
<td>0.741</td>
</tr>
</tbody>
</table>
RQ2B, IMPROVEMENT

WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?

Ferret

<table>
<thead>
<tr>
<th>Energy reduction</th>
<th>Approximation Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>43.19%</td>
<td>0.154</td>
</tr>
<tr>
<td>60.79%</td>
<td>39.873</td>
</tr>
<tr>
<td>75.53%</td>
<td>78.800</td>
</tr>
<tr>
<td>75.55%</td>
<td>1550.200</td>
</tr>
<tr>
<td>76.21%</td>
<td>2669.710</td>
</tr>
<tr>
<td>79.88%</td>
<td>6221.220</td>
</tr>
</tbody>
</table>

Bobby R. Bruce
**RQ2B, IMPROVEMENT**

**WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?**

---

**Ferret**

<table>
<thead>
<tr>
<th>Energy reduction</th>
<th>Approximation Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>43.19%</td>
<td>0.154</td>
</tr>
<tr>
<td>60.79%</td>
<td>39.873</td>
</tr>
<tr>
<td>75.53%</td>
<td>78.800</td>
</tr>
<tr>
<td>75.55%</td>
<td>1550.200</td>
</tr>
<tr>
<td>76.21%</td>
<td>2669.710</td>
</tr>
<tr>
<td>79.88%</td>
<td>6221.220</td>
</tr>
</tbody>
</table>
### RQ2B, IMPROVEMENT

**WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?**

<table>
<thead>
<tr>
<th>Energy reduction</th>
<th>Approximation Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.23%</td>
<td>0.000</td>
</tr>
<tr>
<td>78.45%</td>
<td>0.003</td>
</tr>
<tr>
<td>92.70%</td>
<td>0.637</td>
</tr>
<tr>
<td>95.53%</td>
<td>1.002</td>
</tr>
<tr>
<td>95.60%</td>
<td>1.043</td>
</tr>
</tbody>
</table>

*OMXPlayer*

*Bobby R. Bruce*
RQ2B, IMPROVEMENT

WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?

**OMXPlayer**

<table>
<thead>
<tr>
<th>Energy reduction</th>
<th>Approximation Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.23%</td>
<td>0.000</td>
</tr>
<tr>
<td>78.45%</td>
<td>0.003</td>
</tr>
<tr>
<td>92.70%</td>
<td>0.637</td>
</tr>
<tr>
<td>95.53%</td>
<td>1.002</td>
</tr>
<tr>
<td>95.60%</td>
<td>1.043</td>
</tr>
</tbody>
</table>

Bobby R. Bruce
**RQ2B, IMPROVEMENT**

**WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?**

---

**Distribution of operators**

<table>
<thead>
<tr>
<th>App</th>
<th>Line deletion</th>
<th>Line copy</th>
<th>Line replace</th>
</tr>
</thead>
<tbody>
<tr>
<td>7zip</td>
<td>5</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Ferret</td>
<td>81</td>
<td>7</td>
<td>69</td>
</tr>
<tr>
<td>Bodytrack</td>
<td>44</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>OMXPlayer</td>
<td>8</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td><strong>Percentage</strong></td>
<td><strong>52.9%</strong></td>
<td><strong>4.2%</strong></td>
<td><strong>42.9%</strong></td>
</tr>
</tbody>
</table>

---

Bobby R. Bruce
RQ2B, IMPROVEMENT

WHAT ADDITIONAL ENERGY IMPROVEMENT CAN BE ACHIEVED WHEN USING APPROXIMATE TEST ORACLES IN PLACE OF EXACT TEST ORACLES?

Average impact of the operators

<table>
<thead>
<tr>
<th>App</th>
<th>Line deletion</th>
<th>Line copy</th>
<th>Line replace</th>
</tr>
</thead>
<tbody>
<tr>
<td>7zip</td>
<td>16.60%</td>
<td>0.00%</td>
<td>7.42%</td>
</tr>
<tr>
<td>Ferret</td>
<td>56.45%</td>
<td>64.74%</td>
<td>43.50%</td>
</tr>
<tr>
<td>Bodytrack</td>
<td>8.27%</td>
<td>0.16%</td>
<td>8.31%</td>
</tr>
<tr>
<td>OMXPlayer</td>
<td>57.01%</td>
<td>71.86%</td>
<td>64.92%</td>
</tr>
<tr>
<td>Percentage</td>
<td>34.58%</td>
<td>46.59%</td>
<td>31.04%</td>
</tr>
</tbody>
</table>
RQ2 CONCLUSIONS

➤ When approximation is permitted the number of modifications that reduce energy consumption increased

➤ The modifications are also capable of reducing energy consumption by a greater extent

➤ Some applications produced better Pareto frontiers than others

➤ Copy operations are rarely effective though all have roughly the same impact when they are.
RQ3, SYNERGY

HOW FREQUENTLY DO SYNERGISTIC AND ANTAGONISTIC EFFECTS OCCUR WHEN COMBINING KNOWN EFFECTIVE MODIFICATIONS?
RQ3, SYNERGY

HOW FREQUENTLY DO SYNERGISTIC AND ANTAGONISTIC EFFECTS OCCUR WHEN COMBINING KNOWN EFFECTIVE MODIFICATIONS?
RQ3, SYNERGY

HOW FREQUENTLY DO SYNERGISTIC AND ANTAGONISTIC EFFECTS OCCUR WHEN COMBINING KNOWN EFFECTIVE MODIFICATIONS?

Effective Modification List

Create all possible pairs
RQ3, SYNERGY

HOW FREQUENTLY DO SYNERGISTIC AND ANTAGONISTIC EFFECTS OCCUR WHEN COMBINING KNOWN EFFECTIVE MODIFICATIONS?

Effective Modification List

Create all possible pairs

Select Random 15%

Bobby R. Bruce
RQ3, SYNERGY

HOW FREQUENTLY DO SYNERGISTIC AND ANTAGONISTIC EFFECTS OCCUR WHEN COMBINING KNOWN EFFECTIVE MODIFICATIONS?

1. Create all possible pairs
2. Select Random 15%
3. For each modification pair

Bobby R. Bruce
RQ3, SYNERGY

HOW FREQUENTLY DO SYNERGISTIC AND ANTAGONISTIC EFFECTS OCCUR WHEN COMBINING KNOWN EFFECTIVE MODIFICATIONS?

1. Effective Modification List
2. Create all possible pairs
3. Select Random 15%
4. For each modification pair
5. Measure in Framework

Bobby R. Bruce
How frequently do synergistic and antagonistic effects occur when combining known effective modifications?

1. Effective Modification List
2. Create all possible pairs
3. Select Random 15%
4. For each modification pair
5. Measure in Framework

Aver(Mod1)
Aver(Mod2)
Aver(Mod1 + Mod2)
Aver(Original)
RQ3, SYNERGY

HOW FREQUENTLY DO SYNERGISTIC AND ANTAGONISTIC EFFECTS OCCUR WHEN COMBINING KNOWN EFFECTIVE MODIFICATIONS?

1. Effective Modification List
2. Create all possible pairs
3. Select Random 15%
4. For each modification pair
5. Measure in Framework

- Aver(Mod1)
- Aver(Mod2)
- Aver(Mod1 + Mod2)
- Aver(Original)

Synergy:
- \( r(P\Delta_1) + r(P\Delta_2) \)

Weak Antagonism:
- \( r(P\Delta_1) \)

Antagonism:
- \( r(P\Delta_2) \)

Bobby R. Bruce
RQ3, SYNERGY

HOW FREQUENTLY DO SYNERGISTIC AND ANTAGONISTIC EFFECTS OCCUR WHEN COMBINING KNOWN EFFECTIVE MODIFICATIONS?

1. Effective Modification List
2. Create all possible pairs
3. Select Random 15%
4. For each modification pair
5. Measure in Framework

Aver(Mod1)
Aver(Mod2)
Aver(Mod1 + Mod2)
Aver(Original)

Synergy
Weak Antagonism
Antagonism

Worth combining
Not worth combining

Bobby R. Bruce
RQ3, SYNERGY

HOW FREQUENTLY DO SYNERGISTIC AND ANTAGONISTIC EFFECTS OCCUR WHEN COMBINING KNOWN EFFECTIVE MODIFICATIONS?

<table>
<thead>
<tr>
<th>App</th>
<th>Synergy</th>
<th>Weak Antagonism</th>
<th>Antagonism</th>
</tr>
</thead>
<tbody>
<tr>
<td>7zip</td>
<td>0.9%</td>
<td>60.4%</td>
<td>38.7%</td>
</tr>
<tr>
<td>Ferret</td>
<td>9.2%</td>
<td>48.8%</td>
<td>42.0%</td>
</tr>
<tr>
<td>Bodytrak</td>
<td>35.3%</td>
<td>40.1%</td>
<td>24.6%</td>
</tr>
<tr>
<td>OMXPlayer</td>
<td>2.6%</td>
<td>48.7%</td>
<td>48.7%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>12.0%</strong></td>
<td><strong>49.5%</strong></td>
<td><strong>38.5%</strong></td>
</tr>
</tbody>
</table>
The majority of modifications are worth combining

However, a significant minority exhibit antagonism, meaning they shouldn’t be combined

This means a greedy approach is unlikely to produce an optimal solution; more advanced search is required
ANY QUESTIONS?

Bobby R. Bruce
ANY QUESTIONS?

Bobby R. Bruce