**Genetic Improvement of LLVM Intermediate Representation**

The LLVM project supports many programming languages. Compiling them to device independent intermediate code. LLVM-IR can be optimised. Intermediate code is compiled into device specific machine code.

**What is LLVM?**
- C/C+
- Fortran
- LLVM IR
- PowerPC
- ARM

**What is Genetic Improvement?**
Genetic Improvement uses evolution to modify existing software. Typically GI is applied to human written source code but it can be applied to anything. E.g. C, C++, Java, Java byte code, assembler, even machine code. Non-program software could include comments, documentation and specifications.

**What is Google’s OLC?**
Open Location Code (OLC) plus codes can identify anywhere on Earth. OLC is open source latlngolc.exe 57.101474 -2.242851 9C9V4Q24+HVJXR32

**What is Uber’s H3?**
H3 is another geospatial indexing system. It uses a hexagonal grid that can be (approximatively) subdivided into finer and finer levels. Like OLC it is an open source C industry standard.

**Delete lines of LLVM IR**
By deleting local registers (i.e. set to zero) or forcing conditional branches, IR remains legal and so compiles and runs.

**Pass 1: which lines can be deleted?**
37-63% of mutants do not change output on test cases. These are used by pass 2.

**Pass 2: Use hill climber to join mutants**
Start with fastest ok pass 1 mutant and add others only if they increase speed.

**Fitness Function**
- Is the mutated binary code different
- Does the mutant program run ok
- Are outputs same as unmutated code
- How long does perf say it took

**Results**

<table>
<thead>
<tr>
<th>C file LOC (well)</th>
<th>LLVM IR total mutate change</th>
<th>no output</th>
<th>Mutant</th>
<th>GI-duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLC</td>
<td>4,356 (127)</td>
<td>25%</td>
<td>141</td>
<td>2</td>
</tr>
<tr>
<td>OLC -r 15</td>
<td>4,356 (127)</td>
<td>25%</td>
<td>141</td>
<td>2</td>
</tr>
<tr>
<td>H3</td>
<td>43,570 (1615)</td>
<td>21%</td>
<td>118</td>
<td>5</td>
</tr>
<tr>
<td>H3 -r 15</td>
<td>43,570 (1615)</td>
<td>21%</td>
<td>118</td>
<td>5</td>
</tr>
</tbody>
</table>

Example: H3 mutation 10508%74 saved 872 instructions by causing clang -O3 to remove condition before function doCoords (which must be called).

**Conclusion**
Genetic Improvement on IR in a few minutes or hours gave 0.5% (Google’s OLC) and 2% (Uber’s H3) speed up even on compiler optimised code for two industrial open source C programs.