The Problem

**Parallelization gap** created by lack of parallelism discovery / planning tools

1. Discovery
2. Planning
3. Enabling Transforms
4. Code Generation
5. Runtime Management

A taxonomy of parallelization

Shortcomings of Current Tools

**Problem 1:** High work coverage does not correlate with parallelizability

**Problem 2:** In planning, region structures and parallelization status must be considered

Bridging the Gap - **pyrprof**

1. Follow time-tested *gprof usage model*
2. Profile not only work but also parallelism
3. Leverage *region structure* and parallelization status

**Discovery Stage**

- Build static region graph
- Static region graph is used during planning phase
- Plan
- Instrumented binary
- ![Dynamic Critical Path Calculation](http://parallel.ucsd.edu/pyrprof)
- Instruments source code with calls to pyrprof-polling library functions
  - LLVM used to inline and highly optimize instrumented code
- pyrprof-lib supports tracking of control and data dependencies
- Data dep: use shadow memory
- Control dep: tracked with special stack
- Parallelism = Work / Critical Path
- Redundant profiling data compressed using a dictionary-based technique

**Planning Stage**

Goal: provide an ordered list of regions ranked by the impact of their parallelization

Example Static Region Graph

- Estimated execution time of the input program
  - *time, profile, P, E*
  - = estimated execution time of the input program
  - *- region to parallelize: profile data from discovery stage
  - P: set of parallelized regions
  - E: set of excluded regions

Execution Time Estimation Model

**Case Study – mpeg encoder (ALPBench)**

Example Iterative Workflow

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Excluded Regions</th>
<th>Top 5 PAR</th>
<th>Confirmed PAR</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>()</td>
<td>A, B, D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(E)</td>
<td>A, B, D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(D, E)</td>
<td>A, B, G</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Planning Effectiveness Comparison**

<table>
<thead>
<tr>
<th>Region</th>
<th>gprof</th>
<th>pyrprof</th>
<th>pyrprof interactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>27</td>
<td>41</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>31</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>E</td>
<td>35</td>
<td>35</td>
<td>5</td>
</tr>
<tr>
<td>F</td>
<td>42</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>G</td>
<td>8</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>H</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>I</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

**Parallelization Plan**

**Availability**

pyrprof is available for free download at: [http://parallel.ucsd.edu/pyrprof](http://parallel.ucsd.edu/pyrprof)