Parameterized Loop Fusion for Automatic Performance Tuning

Yuan Zhao Qing Yi Ken Kennedy Dan Quinlan Richard Vuduc
Rice University Lawrence Livermore National Laboratory

Context: Automatic Performance Tuning

- has been used to tune performance of libraries and programs, e.g. ATLAS, FFTW
- uses profiling information (e.g. running time) as a guidance to search through a space of optimization configurations
- tunes a software to a specific architecture (instruction set, number of registers, cache size/associativity, etc) for the best performance

Focus of Study: Loop Fusion

- example:

```
for(i = 0; i < size; i++)
  a[i] = c[i] + d;
for(i = 0; i < size; i++)
  b[i] = c[i] – d;
```

```
for(i = 0; i < size; i++) {
  a[i] = c[i] + d;
  b[i] = c[i] – d;
}
```

- benefits: less loop iterating cost, closer memory reuses (less memory traffic, less cache miss)
- problem: fusing all fusible loops together (maximal fusion) may not yield the best performance due to the limited resources on the architecture
- goal: search for the loop fusion result with the best performance

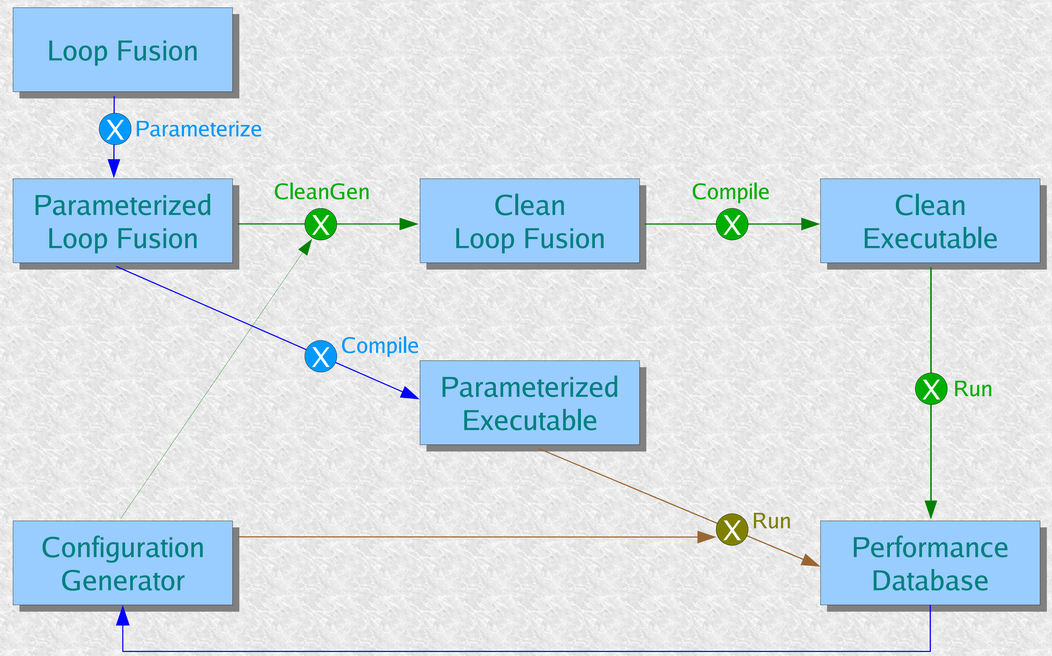
Our Method: Parameterized Loop Fusion

- example:

```
VN = DynamicFusionBegin(1, 2, &V1, &V2);
for(V = 1; V <= VN; V++)
  for(i = 0; i < size; i++) {
    if(V == V1)
      a[i] = c[i] + d;
    if(V == V2)
      b[i] = c[i] – d;
  }
DynamicFusionEnd(1);
```

- parameters: V1, V2, VN
- configuration: an instantiation of (V1, V2, VN)
  e.g. V1 = V2 = VN = 1, two loops are fused
  V1 = 1, V2 = VN = 2, two loops not fused
- number of configurations
  in-order: (d+1)^(n-1), d: loop nest depth, n: number of statements
  out-of-order: sterling number if d = 1

Tuning Process



Research in Progress

- build the tuning framework upon the ROSE framework developed here at LLNL
- verify the performance correlation between parameterized version and clean version
- study the properties of the search space of loop fusion
- study the effectiveness of various search heuristics

Acknowledgments

- This research is a part of the ongoing collaboration among Rice University, Lawrence Livermore National Laboratory, and Jack Dongarra's group at University of Tennessee.
- This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.