CS 526
Advanced Compiler Construction

http://misailo.cs.illinois.edu/courses/cs526
DEPENDENCE TRANSFORMS

The slides adapted from Vikram Adve
Reordering Transformation

**Definition.** Legal Transformation preserves the meaning of that program, i.e., all externally visible outputs are identical to the original program, and in identical order.

- We consider two programs equivalent (i.e., the transformation preserving the program meaning) if on the same inputs both the original and transformed programs, after being executed, produce the same outputs.

**Theorem.** A **reordering** transformation that preserves all data dependences in a program is a **legal** transformation.

- See Lecture 6 for an argument why.
Motivation

**Memory hierarchy optimizations**
Goal 1: Improving reuse of data values within loop nest
Goal 2: Exploit reuse to reduce cache, TLB misses

**Tiling**
Goal 1: Exploit temporal reuse when data size > cache size
Goal 2: In parallel loops, reduce synchronization overhead

**Software Prefetching**
Goal: Prefetch predictable accesses k iterations ahead

**Software Pipelining**
Goal: Extract ILP from multiple consecutive iterations

**Automatic parallelization** Also, auto-vectorization
Goal 1: Enhance parallelism
Goal 2: Convert scalar loop to explicitly parallel
Goal 3: Improve performance of parallel code
# Reordering Transformations

<table>
<thead>
<tr>
<th>Name</th>
<th>Purpose</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Preprocessing transformations</strong></td>
<td></td>
</tr>
<tr>
<td>Loop normalization</td>
<td>Make loops canonical</td>
<td>Simplify, improve dep. analysis</td>
</tr>
<tr>
<td>Ind. var. substitution</td>
<td>Identify aux. induction vars</td>
<td>Improve dependence information</td>
</tr>
<tr>
<td>Scalar expansion</td>
<td>Replace scalar with array</td>
<td>Eliminate spurious dependences</td>
</tr>
<tr>
<td>Scalar/array privatization</td>
<td>Treat var. as iteration-private</td>
<td>Eliminate spurious dependences</td>
</tr>
<tr>
<td>Variable renaming</td>
<td>Use multiple copies of vars</td>
<td>Eliminate anti- and output-dependences</td>
</tr>
<tr>
<td>Reduction recognition</td>
<td>Recognize reductions</td>
<td>Ignore special-case dependences</td>
</tr>
<tr>
<td></td>
<td><strong>Reordering transformations</strong></td>
<td></td>
</tr>
<tr>
<td>Loop interchange</td>
<td>Change loop nesting order</td>
<td>Cache, parallelism, vectorization</td>
</tr>
<tr>
<td>Loop strip-mining</td>
<td>Make 2 nested loops</td>
<td>&quot;</td>
</tr>
<tr>
<td>Loop skewing</td>
<td>Change wavefront loop to parallel</td>
<td>Improve loop parallelism</td>
</tr>
<tr>
<td>Loop reversal</td>
<td>Run loop backwards</td>
<td>Reduce array storage</td>
</tr>
<tr>
<td>Index set splitting</td>
<td>Break loop by index space</td>
<td>Remove some deps.</td>
</tr>
<tr>
<td>Loop distribution</td>
<td>Break loop by statements</td>
<td>Simplify parallelization, vectorization</td>
</tr>
<tr>
<td>Loop alignment</td>
<td>Change carried to indep.</td>
<td>Simplify parallelization, vectorization</td>
</tr>
<tr>
<td>Loop fusion</td>
<td>Join loops by statements</td>
<td>Improve cache reuse</td>
</tr>
</tbody>
</table>
Loop Interchange

**Informal Definition:** Change nesting order of loops in a **perfect loop nest**, with no other changes.

```plaintext
for i=2 to N
    for j=2 to M-1
    end for
end for

for j=2 to M-1
    for i=2 to N
    end for
end for
```
Uses of Loop Interchange

1. Move independent loop innermost
2. Move independent loop outermost
3. Make accesses stride-1 in memory
4. Loop tiling (combine with strip-mining)
5. Unroll-and-jam (combine with unrolling)
Loop Interchange

Direction Vectors and Loop Interchange:
If $\delta$ is a direction vector of a particular dependence $S_1 \rightarrow S_2$ in a loop nest and the order of loops in the loop nest is permuted, then the same permutation can be applied to $\delta$ to obtain the new direction vector for the conflicting instances of $S_1$ and $S_2$.

**Direction Matrix:** A matrix where each row is the direction vector of a single dependence, i.e.,
each row $\leftrightarrow$ a dependence
each column $\leftrightarrow$ a loop
Direction Matrix

Direction Matrix:
each row ↔ a dependence
each column ↔ a loop

for i = 2 to N
    for j = 2 to M-1
        Sp: A[i,j] = ... * B[i-1,j-1]
    endfor
endfor

Sp→Sq: A[i,j]/A[i,j] = =
Sp→Sq: A[i,j]/A[i-1,j] + =
Sq→Sp: B[i,j]/B[i-1,j-1] + +
Direction Matrix (Illegal)

Direction Matrix:
each row ↔ a dependence
each column ↔ a loop

for $i = 2$ to $N$
    for $j = 2$ to $M-1$
        Sp: $A[i,j] = \ldots \times B[i-1,j-1]$
    endfor
endfor

\[
\begin{align*}
\text{Sp} \rightarrow \text{Sq}: A[i,j]/A[i,j] &= = \\
\text{Sp} \rightarrow \text{Sq}: A[i,j]/A[i-1,j+1] &= + - \\
\text{Sq} \rightarrow \text{Sp}: B[i,j]/B[i-1,j-1] &= + +
\end{align*}
\]
Loop Interchange Properties

Legality: A permutation of the loops in a perfect nest is legal iff the direction matrix, after the permutation is applied, has no “-” direction as the leftmost non-“=” direction in any row.

Intuition: To preserve the dependencies, consider the cases before transformation of (=,=) [independent], (=,+), and (+,=) [the dependence is still carried but by the outer (resp. inner loops)], (+,+) [Dependence is still carried]. But (+,-) is illegal since the antidependence turns into a true dependence.

In other words, the vector after transformation should be lexicographically greater than “0” vector.

E.g., (1, -1) > (0,0) but (-1, 1) < (0,0)
Loop Interchange Properties

**Legality:** A permutation of the loops in a perfect nest is legal iff the direction matrix, after the permutation is applied, has no “-” direction as the leftmost non-“=“ direction in any row.

**Profitability:** machine-dependent:
1. vector machines
2. parallel machines
3. caches with single outstanding loads
4. caches with multiple outstanding loads
Applying Loop Interchange

1. **Single ′+′ entry:** a “serial loop”
   - Move loop outermost for vectorization
   - Move loop innermost for parallelization

2. **Multiple ′+′ entries:** Outermost one carries dependence
   - Loop carrying the dependence *changes* after permutation!
   - May still benefit by moving carried-dependences to the outermost loop
Example

for i = 1 to n
    for j = 1 to m
    end for
end for

for i = 1 to n
    for j = 1 to m // vectorize
        A[i+1, j] = A[i, j]
        + B[i, j]
    end for
end for

parallel for j = 1 to m
    for i = 1 to n
        A[i+1, j] = A[i, j]
        + B[i, j]
    end for
end for
Loop Reversal

**Informal Definition:** Reverse the order of execution of the iterations of a loop

```
for i=2 to N
    for j=2 to M-1
        for k=1 to L
        endfor
    endfor
endfor
```
Loop Reversal

for i=2 to N
  for j=2 to M-1
    for k=1 to L
    endfor
  endfor
endfor

for i=2 to N
  for j=2 to M-1
    for k=L to 1 step -1
    endfor
  endfor
endfor

= + -
+ = -

= ++
+ = +
Legality of Loop Reversal

The loop that is reversed should not carry the dependence

Recall, *Legality*: the vector after transformation should be lexicographically greater than “0” vector.

E.g., \((1, -1) \succ (0,0)\) but \((-1, 1) \prec (0,0)\)
Uses of Loop Reversal

Convert a ’-’ to a ’+’ in a direction vector to enable other transformations, e.g., loop interchange.

Scalarize a vector statement (e.g., in Fortran 90) by ensuring that values are read before being written.


• Scalarized code:
  
  ```
  for i = 64 to 2 step -1
    A[i] = A[i-1] \times e
  endfor
  ```
Loop Distribution

**Informal Definition:** Convert a loop nest containing two or more statements into two or more distinct loop nests so that each statement appears in only a single resulting loop nest.

```plaintext
for i = 2 to N
  S1:    A[i] = B[i] + C[i]
  S2:    D[i] = A[i] * 2.0
end for
```

```plaintext
for i = 2 to N
  S1:    A[i] = B[i] + C[i]
end for
```

```plaintext
for i = 2 to N
  S2:    D[i] = A[i] * 2.0
end for
```
Loop Distribution Applications

- Create perfect loops nests for other transformations like loop interchange
- Convert a loop-carried dependence within a loop into a loop-independent dependence crossing two loops:

```plaintext
for i=2 to N
  S1:      A[i] = B[i] + C[i]
  S2:      D[i] = A[i-1] * 2.0
end for
```

```plaintext
for i=2 to N
  S1:      A[i] = B[i] + C[i]
  S2:      D[i] = A[i-1] * 2.0
end for
```
Maximal Loop Distribution

• Identify the SCCs of the data dependence graph, to group statements in an SCC in a single loop nest
• Sort the SCCs using a topological sort on the dependence graph
• Generate distinct loop nests, one for each SCC, in sorted order

Reminder:
• **Strongly connected graph**: a directed graph in which there is a path between all pairs of vertices.
• **Strongly connected component (SCC)** is a maximal strongly connected subgraph
Loop Fusion

**Informal Definition:** Merge two or more distinct (perhaps non-adjacent) loops with identical loop bounds into a single loop.

```plaintext
for i=1 to N
    A[i] = i*i
end for
for i=1 to N
    B[i] = A[i] + 1
end for
```
Loop Fusion

for i=1 to M
    for j=1,N-1
        A[j,i] = i*i + j*j
    end for

    for j=1 to N
        B[j,i] = A[j,i] + i + j
    end for
end for

// peel last iteration:
for j=1 to N-1
    A[j,i] = i*i + j*j
    B[j,i] = A[j,i] + i + j
end for

for i=1 to M
    for j=1 to N-1
        A[j,i] = i*i + j*j
        B[j,i] = A[j,i] + i + j
    end for
end for
Loop Fusion Motivation

• Increase cache reuse (if same array accessed in two loops) Fundamental optimization for array languages (e.g., Fortran 90, HPF, MATLAB, APL)

Example in F90:

\[
\]

• Increase granularity of parallelism (work per iteration) Important for shared-memory parallelism (the model with parallel loop and barriers)
Legality of Loop Fusion

Fusion-Preventing Dependence: A loop-independent dependence from S1 to S2 in different loops is fusion-preventing if fusing the two loops causes the dependence to become a loop-carried dependence from S2 to S1.

Legality of Loop Fusion: Two loops can be fused if all three conditions are satisfied:

1. Both have identical bounds (transform loops if needed)
2. There is no fusion-preventing dependence between them.
3. There is no path of loop-independent dependences between them that contains a loop or statement that is not being fused with them.
Loop Fusion: Illegal Cases

for i=1 to M
    for j=2 to N
        A[j,i] = B[j-1,i] * 2
    end for
end for

for j=2 to N
end for

for i=1 to M
    for j=2 to N
        t[j] = B[j-1,i]
    end for
end for

for j=2 to N
    A[j,i] = t[j] * 2
end for

for j=2 to N
end for

Create temporary array to make fusion possible
Loop Alignment

**Informal Definition:** Eliminate a carried dependence by increasing the number of iterations and executing statements on different subsets of the iterations

*(Always safe)*

for i=2 to N
    A[i] = B[i] + C[i]
    D[i] = A[i-1] * 2.0
end for

\[
i = 1
D[i+1] = A[i] \times 2
\]

for i=2 to N-1
    A[i] = B[i] + C[i]
    D[i+1] = A[i] \times 2.0
end for

\[
i = N
A[i] = B[i] + C[i]
\]
Loop Strip Mining

**Informal Definition** Convert a single loop into two nested loops for a specified “block size”

(*Always safe.*)

```plaintext
for i=1 to N
    A[i] = x + B[i] * 2
end for

for ii=1 to N step B
    for i=ii to min(ii+B-1, N)
        A[i] = x + B[i] * 2
    end for
end for
```
Loop Strip Mining Applications

- **Loop tiling:** strip-mine and then interchange multiple uses. Can be useful for increasing cache locality or blocking parallel loops;

- **Prefetching:** strip-mine by cache line size; prefetch once per outer iteration

- **Instruction scheduling:** strip-mine and then unroll inner loop
Tiling Example

for $II_1' := 0$ to $5$ by $2$ do
  for $II_2' := 0$ to $11$ by $2$ do
    for $l_1'$ := $II_1'$ to min(5,$II_1'$ + 1) do
      for $l_2'$ := max($l_1'$, $II_2'$) to min(6+$l_1'$, $II_2'$+1) do
        $a[l_2' + 1] := 1/3 * (a[l_2'] + a[l_2' + 1] + a[l_2' + 2]);$

Fig. 4. Iteration space and dependences of tiled code from Fig. 1(b).
More details:

Optimizing Compilers for Modern Architectures

Allen and Kennedy

Academic Press
Polyhedral Compilation

Brief Introduction to Polyhedral Compilation Techniques:
Basic polyhedral concepts in program analysis
Iteration spaces; array references
Dependence analysis
Loop transformations: representation
Loop transformations: code generation
Polyhedra

**k-tuple:** A point in $\mathbb{Z}^k$, e.g., $(1, -4, 3)$ or $J = (i_1, i_2, \ldots, i_k)$

**Tuple set:** A set of tuple points $(0, 1, 2), (2, 3) \ldots$

**Tagged tuple set:** A set of tuple points $A(1, 2), C(3)$
- Can be represented as a tuple, where e.g., $\text{map}(A) = 0, \text{map}(C) = 2$

**Polyhedron:** A tuple set defined by affine inequalities

**General:** $\{(i_1, i_2, \ldots, i_k) : A \cdot \vec{i} \leq \vec{U}\}$
- e.g. $\{(i_1, i_2) : L_1 \leq i_1 < U_1 \land L_2 \leq i_2 < U_2\}$
- Focus on convex polyhedral
- Integer polyhedron: all in/out points are integers
- Integer hull: set of integer points that bounds rational polyhedron
Tuple Relations

A tuple relation (or relation or mapping) is a mapping from tuple sets to tuple sets, e.g.,
\[
\{(i,j) \rightarrow (ii,jj) : 0 \leq i < N \land 0 \leq j < N \land ii = i \land jj = i + j - 1\}
\]

A relation, \(R\), “applied” to a tuple set, \(S\), yields a new tuple set, \(R(S)\).
E.g., \(S = \{(i) : 0 \leq i \leq N\}\), \(R = \{(i) \rightarrow (ii) : 0 \leq i \leq N \land ii = 2i + 1\}\),
results in \(R(S) = \{(ii) : \exists k : ii = 2k + 1 \land 1 \leq ii \leq 2N + 1\}\).
Dependence Analysis Steps

1. Extract model from the code
   - Affine iteration spaces as Polyhedra
   - Array references as polyhedral mappings

2. Dependence analysis:
   - Turn into polyhedral satisfaction problem

3. Transformations:
   - Permutations/transformations on the model, specified by tuple relations
   - Generate code from the model (original code and the transformed iteration spaces)
Affine Iteration Spaces as Polyhedra

Every statement in the program has an associated iteration space, describing the enclosing loops:

\[ L = \{(i_1, i_2, \ldots i_k) : L_1 \leq i_1 < U_1 \]
\[ \land L_2 \leq i_2 < U_2 \]
\[ \land L_k \leq i_k < U_k \} \]

- For polyhedral analysis, \( L_i, U_i \) must be affine functions of index variables (i), loop-invariant program variables and constants.
Array References as Polyhedral Mappings

for $i_1 = L_1$ to $U_1$
    $S_1$
    for $i_2 = L_2$ to $U_2$
        $S_2$
        . . .
        for $i_k = L_k$ to $U_k$
            $A[i_1,\ldots,i_k] = \ldots$
            end for
        . . .
    end for
end for

Every array reference in the program is a mapping from the iteration space (of the statement) to array elements. E.g.,

$$L \rightarrow A : \{(\vec{i}, \vec{a}) : \vec{i} \in L \land a_1 = f_1(\vec{i}) \ldots \land a_r = f_r(\vec{i})\}$$

- For polyhedral analysis, $f_i$, must be affine functions of index variables ($i$), loop-invariant program variables and constants.
Checking for Data Dependence

There is a data dependence between
\[ A(f_1(\vec{i}), f_2(\vec{i}), \ldots, f_r(\vec{i})) \] and \[ A(g_1(\vec{i}), g_2(\vec{i}), \ldots, g_r(\vec{i})) \]

iff the following polyhedron contains integer points:

\[ \{(i_1, i_2, \ldots, i_r, j_1, j_2, \ldots, j_r) : \vec{i} \in L \land \vec{j} \in L \land f_1(\vec{i}) = g_1(\vec{j}) \land \ldots \land f_r(\vec{i}) = g_r(\vec{j})\} \]
Program Transformations

Program transformations as polyhedral mappings: Many program transformations can be represented as a mapping (for each original program statement) from its iteration space in the original program to its iteration space in the transformed program.

Loop reordering transformations: a transformation on a perfect loop nest that reorders the loop iteration space but does not modify the relative order of statements within the innermost loop (sometimes called an atomic block).

\[ L \rightarrow L : \{(i) \rightarrow (\vec{i}) : \vec{i} \in L \wedge ii_1 = \varphi_1(i) \wedge \ldots \wedge ii_k = \varphi_k(i)\} \]
Loop Transformations and Matrices

Alternate representation for loop transformations – as a matrix:
\[ \Phi(i) = T \cdot \tilde{i} + \tilde{t} \]

- The transformation is affine iff \( T \) is a constant matrix and \( \tilde{t} \) is a parametric vector consisting of loop-invariant program variables and constants.
- Each column in the matrix product represents a single input loop. Each row in the matrix product represents a single output loop.
- The transformation is called *unimodular* if \( T \) is unimodular (i.e., square integer matrix with determinant +1 or -1)
Loop Transformations and Matrices

A transformation is called *unimodular* if the matrix $T$ is unimodular (i.e., square integer matrix with determinant +1 or -1)

$$T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tilde{t} = \bar{0}$$

Loop interchange:

Loop reversal: $T = [-1], \tilde{t} = (U_1 - 1)$

Legality of the transformation: $T \cdot \tilde{t} > 0$
Example Transformations

Loop reversal: $\Phi = \{(i) \rightarrow (ii) : L_1 \leq i \leq U_1 \land ii = U_1 - i + 1\}$

\[
\begin{align*}
\text{do } & i = L_1 \text{ to } U_1 \\
\quad & A(i) = B(i) + C(i) \\
\text{enddo} & \quad \Rightarrow \quad \text{do } ii = U_1 \text{ to } L_1 \text{ by } -1 \\
\quad & A(ii) = B(ii) + C(ii) \\
\text{enddo}
\end{align*}
\]
Example Transformations

Loop reversal: \( \Phi = \{(i) \rightarrow (ii) : L_1 \leq i \leq U_1 \land ii = U_1 - i + 1\} \)

\[
\begin{align*}
do & \ i = L\_1 \ to \ U\_1 \\
& \quad \ A(i) = B(i) + C(i) \\
& \quad \ enddo  \\
\implies & \ do \ ii = U\_1 \ to \ L\_1 \ by \ -1 \\
& \quad \ A(ii) = B(ii) + C(ii) \\
& \quad \ enddo
\end{align*}
\]

Loop interchange: \( \Phi = \{(i,j) \rightarrow (jj,ii) : L_1 \leq i \leq U_1 \land L_2 \leq j \leq U_2 \land ii = i \land jj = j\} \)

\[
\begin{align*}
do & \ i = L\_1 \ to \ U\_1 \\
& \quad \ do \ j = L\_2 \ to \ U\_2 \\
& \quad \quad \ A(i,j) = B(i+j,i-j) + 1 \\
& \quad \quad \ enddo \\
& \quad \ enddo
\end{align*}
\]

\[
\begin{align*}
do & \ jj = L\_2 \ to \ U\_2 \\
& \quad \ do \ ii = L\_1 \ to \ U\_1 \\
& \quad \quad \ A(ii,jj) = B(ii+jj,ii-jj) + 1 \\
& \quad \quad \ enddo \\
& \quad \ enddo
\end{align*}
\]
Imperfect Loop Nests

**General approach:** Add an extra ("sequencing") dimension in the iteration space to enforce ordering on individual statements:

```plaintext
for i = L_1 to U_1
    S1(i)
end for

for j = L_2 to U_2
    S2(i,j)
end for

S3(i)

L(S1) = \{(i, 0, j): L1 \leq i \leq U1 \land j = L2\}
L(S2) = \{(i, 1, j): L1 \leq i \leq U1 \land L2 \leq j \leq U2\}
L(S3) = \{(i, 2, j): L1 \leq i \leq U1 \land j = U2\}
```
Pros and Cons

Pros:
• Principled representation
• Fine-grained optimization and analysis using mathematical programming
• Simplify loop transformations

Cons:
• In general, NP-complete problem: boils forwn to Integer programming
• Memory consuming, especially for irregular nests with control flow
References

Courses/Lectures:
• Louis-Noël Pouchet course: http://web.cse.ohio-state.edu/~pouchet/#lectures
• Pollylabs video and written tutorials: http://www.pollylabs.org/education.html

Tools: GCC Graphite, URUK, Omega, Loop…

Polly (LLVM):
• Tool: http://polly.llvm.org
• Interactive playground: http://playground.pollylabs.org/
We will analyze the Successive Over Relaxation (SOR) kernel for solving partial differential equations:

\[
t = 0;
\]

\[
\text{for } (i=1; i < N-1; i++)
\]

\[
\text{for } (j=1; j < N-1; j++) \{
\]

\[
\]

\[
S2: \quad A[i,j] = 0.25 \times t;
\]

\[
}\}
\]

19. **[5 points]**: Construct the dependence graph for the code fragment below. Mark each edge as a true, anti, or output dependence in the usual way. (Do not show input dependences.) Label each dependence edge with its direction vector. If a particular direction vector entry has multiple values (e.g., + and =), show that as multiple dependence edges with separate direction vectors.

20. **[5 points]**: Suggest a transformation that will reduce the number of dependences. Write down the transformed loop nest and its dependence graph.

Remove the temporary variable \( t \), since its value is used only once. We remove antidependence and output dependence on the temporary variable (recall the discussion in class).