# Enhancing FineGrained Parallelism 

Loop vectorization,
Loop distribution,
Scalar expansion
Scalar and array renaming

## Fine-Grained Parallelism

- Theorem 2.8. A sequential loop can be converted to a parallel loop if the loop carries no dependence.
- Fine-grained parallelism (vectorization)
- Want to convert loops like:

```
DO I=1,N
        X(I) = X(I) +C
ENDDO
to X(1:N) = X(1:N) + C (Fortran 77 to Fortran 90)
```

- However:

```
DO I=1,N
is not equivalent to \(X(2: N+1)=X(1: N)+C\)
        \(X(\mathrm{I}+1)=\mathrm{X}(\mathrm{I})+\mathrm{C}\)
ENDDO
```

- Techniques to enhance fine-grained parallelism
- Goal: make more inside loops parallelizable
- Transform loops: Loop distribution, loop interchange
- Transform data: scalar Expansion, scalar and array renaming


## Loop Distribution

- Can dependence-carrying loops be vectorized?

$$
\begin{aligned}
& \text { DO I = 1, N } \\
& \text { S1 } A(I+1)=B(I)+C \\
& \text { S2 } D(I)=A(I)+E \\
& \text { ENDDO }
\end{aligned}
$$

- Safety of loop distribution
- There must be no dependence cycle connecting statements in different loops after distribution
DO I = 1, N
S1 $\quad \mathrm{A}(\mathrm{I}+1)=\mathrm{B}(\mathrm{I})+\mathrm{C}$
S2 $\quad B(I+1)=A(I)+E$
ENDDO


## Loop Interchange

- Most statements are surrounds by more than one loops

$$
\begin{aligned}
& \mathrm{DO} \mathrm{I}=1, \mathrm{~N} \\
& \mathrm{DO} \mathrm{~J}=1, \mathrm{M} \\
& \mathrm{~S} 1 \quad \mathrm{~A}(\mathrm{I}+1, \mathrm{~J})=\mathrm{A}(\mathrm{I}, \mathrm{~J})+\mathrm{B} \\
& \text { ENDDO } \\
& \text { ENDDO }
\end{aligned}
$$

- Dependence from S1 to itself carried by outer loop
- Inner loop can be parallelized

DO I =1, N
S1 $A(I+1,1: M)=A(I, 1: M)+B$
ENDDO

- Loop interchange: change the nesting order of loops


## Applying Loop Distribution

- procedure codegen(R, k, D);
$R$ :code to transform; $k$ : the loop level to optimize;
D: dependence graph for $R$
- Find strongly-connected regions $\{\mathrm{S} 1, \mathrm{~S} 2, \ldots, \mathrm{Sm}\}$ of D ;
- $\mathrm{Rp}=$ reduce each Si to a single node in R

Dp = the dependence graph of Rp

- For each node pi in topological order of nodes in Dp
- Let Di be the dependence graph of pi at loop level $k+1$;
$\square$ if Di is cyclic then
- generate a level-k DO statement;
- codegen (pi, k+1, Di);
- generate the level-k ENDDO statement;
$\square$ else
- Try to vectorize inner loops in pi


## Loop Distribution and Vectorization

DO I = 1, 100
$S_{1} X(I)=Y(I)+10$
DO $J=1,100$
$\mathrm{S}_{2} \quad \mathrm{~B}(\mathrm{~J})=\mathrm{A}(\mathrm{J}, \mathrm{N})$

$$
\mathrm{DO} \mathrm{~K}=1,100
$$

$\mathrm{S}_{3} \quad \mathrm{~A}(\mathrm{~J}+1, \mathrm{~K})=\mathrm{B}(\mathrm{J})+\mathrm{C}(\mathrm{J}, \mathrm{K})$ ENDDO
$S_{4} \quad Y(I+J)=A(J+1, N)$ ENDDO
ENDDO


## Loop Distribution and Vectorization

- codegen $\left.\left(\left\{\mathrm{S}_{2}, \mathrm{~S}_{3}, \mathrm{~S}_{4}\right\}, 2\right\}\right)$
- level-1 dependences are stripped off

$$
\begin{aligned}
& \text { DO } I=1,100 \\
& \text { DO } J=1,100 \\
&\text { codegen } \left.\left(\left\{S_{2}, S_{3}\right\}, 3\right\}\right)
\end{aligned}
$$

## ENDDO

$S_{4} \quad Y(I+1: I+100)=A(2: 101, N)$ ENDDO
$X(1: 100)=Y(1: 100)+10$


## Loop Distribution and Vectorization

- codegen ( $\left.\left\{\mathrm{S}_{2}, \mathrm{~S}_{3}\right\}, 3\right\}$ )
- level-2 dependences are stripped off

$$
\begin{aligned}
& \text { DO } I=1,100 \\
& \text { DO } J=1,100 \\
& \text { B (J) }=A(J, N) \\
& A(J+1,1: 100)=B(J)+C(J, 1: 100)
\end{aligned}
$$

ENDDO
$\mathrm{Y}(\mathrm{I}+1: \mathrm{I}+100)=\mathrm{A}(2: 101, \mathrm{~N})$
ENDDO

$$
X(1: 100)=Y(1: 100)+10
$$

```
DO I = 1, 100
\(S_{1} \quad X(I)=Y(I)+10\)
    DO J = 1, 100
        \(\mathrm{B}(\mathrm{J})=\mathrm{A}(\mathrm{J}, \mathrm{N})\)
        DO K = 1, 100
            A \((\mathrm{J}+1, \mathrm{~K})=\mathrm{B}(\mathrm{J})+\mathrm{C}(\mathrm{J}, \mathrm{K})\)
        ENDDO
    \(\mathrm{S}_{4} \quad \mathrm{Y}(\mathrm{I}+\mathrm{J})=\mathrm{A}(\mathrm{J}+1, \mathrm{~N})\)
    ENDDO
```

ENDDO


## Loop Interchange

- A reordering transformation that
- Changes the nesting order of loops
- Example

```
DO I = 1, N
        DO J = 1,M
    S A(I,J+1) = A(I,J) + B - Direction vector: ( }=,<
        ENDDO
    ENDD
```

- After loop interchange

DO J = 1, M
DO $I=1, N$
S $\quad A(I, J+1)=A(I, J)+B$

- Direction vector: $(<,=)$

ENDDO

- Leads to

```
DO J = 1,M
S A(1:N,J+1) = A(1:N,J) + B
ENDDO
```


## Safety of Loop Interchange

- Not all loop interchanges are safe

$$
\begin{aligned}
& \text { DO } \mathrm{J}=1, \mathrm{M} \\
& \text { DO } \mathrm{I}=1, \mathrm{~N} \\
& \quad \mathrm{~A}(\mathrm{I}, \mathrm{~J}+1)=\mathrm{A}(\mathrm{I}+1, \mathrm{~J})+\mathrm{B} \quad \text { Direction vector: }(<,>) \\
& \text { ENDDO } \\
& \text { ENDDO }
\end{aligned}
$$



## Loop Interchange: Safety

- Direction matrix of a loop nest contains
- A row for each dependence direction vector between statements contained in the nest.

```
DO I = 1, N
    DO J = 1,M
                DO K = 1,L
                A(I+1,J+1,K) = A(I,J,K) + A(I,J+1,K+1)
                        ENDDO
ENDDO
ENDDO
```

$$
\left(\begin{array}{lll}
< & < & = \\
< & = & >
\end{array}\right]
$$

- Theorem 5.2 A permutation of the loops in a perfect nest is legal if and only if
- the direction matrix, after the same permutation is applied to its columns, has no ">" direction as the leftmost non-"=" direction in any row.


## Loop Interchange: Profitability

- Profitability depends on architecture

$$
\begin{aligned}
& \text { DO } I=1, N \\
& \text { DO } J=1, M \\
& \text { DO } K=1, L \\
& S \quad A(I+1, J+1, K)=A(I, J, K)+B
\end{aligned}
$$

- For SIMD machines with large number of FU's:

$$
\begin{aligned}
& \text { DO } I=1, N \\
& S \quad A(I+1,2: M+1,1: L)=A(I, 1: M, 1: L)+B
\end{aligned}
$$

- For Vector machines: vectorize loops with stride-one access

```
DO J = 1, M
        DO K = 1, L
```

S

$$
A(2: N+1, J+1, K)=A(1: N, J, K)+B
$$

- For MIMD machines with vector execution units: cut down synchronization costs

$$
\begin{aligned}
& \text { PARALLEL DO } K=1, L \\
& \text { DO } J=1, M \\
& A(2: N+1, J+1, K)=A(1: N, J, K)+B
\end{aligned}
$$

## Loop Shifting

- Goal: move loops to "optimal" nesting levels
- Apply loop interchange repeatedly when safe
- Theorem 5.3 In a perfect loop nest, if loops at level $\mathrm{i}, \mathrm{i}+1, \ldots, \mathrm{i}+\mathrm{n}$ carry no dependence, it is always legal to shift these loops inside of loop $\mathrm{i}+\mathrm{n}+1$. Furthermore, these loops will not carry any dependences in their new position.


## Loop Selection

- Consider:

```
        DO I = 1, N
        DO J = 1, M
    S \(\quad \mathrm{A}(\mathrm{I}+1, \mathrm{~J}+1)=\mathrm{A}(\mathrm{I}, \mathrm{J})+\mathrm{A}(\mathrm{I}+1, \mathrm{~J})\)
        ENDDO
    ENDDO
    - Direction matrix: \(\quad(=<)\)
```

- Interchanging the loops can lead to:
DO J = 1, M
$A(2: N+1, J+1)=A(1: N, J)+A(2: N+1, J)$
ENDDO
- Which loop to shift?
- Select a loop at nesting level $p \geq k$ that can be safely moved outward to level $k$ and shift the loops at level $k, k+1, \ldots, p-1$ inside it


## Heuristics for selecting loop level

- Goal: maximize \# of parallel loops inside
- If the level-k loop carries no dependence,
$\square$ let $p$ be the level of the outermost loop that carries a dependence
- If the level-k loop carries a dependence,
- let $p$ be the outermost loop that can be safely shifted outward to position k and that carries a dependence direction vector $d$ which has " $=$ " in every position but the $p^{\text {th }}$. If no such loop exists, let $p=k$.



## Loop Shifting Example

$$
\begin{aligned}
& \text { DO } I=1, N \\
& \text { DO J }=1, N \\
& \text { DO K }=1, N \\
& S \quad A(I, J)=A(I, J)+B(I, K) * C(K, J)
\end{aligned}
$$

- S has true, anti and output dependences on itself
- Vectorization fails as recurrence exists at innermost level
- Use loop shifting to move K-loop to the outermost

$$
\begin{aligned}
& \mathrm{DO} K=1, N \\
& \text { DO } \mathrm{I}=1, N \\
& \mathrm{DO} \mathrm{~J}=1, N \\
& \mathrm{~S} \quad \mathrm{~A}(\mathrm{I}, \mathrm{~J})=\mathrm{A}(\mathrm{I}, \mathrm{~J})+\mathrm{B}(\mathrm{I}, \mathrm{~K}) * \mathrm{C}(\mathrm{~K}, \mathrm{~J})
\end{aligned}
$$

- Parallelization is now possible

$$
\begin{aligned}
& \text { DO } K=1, N \\
& \text { FORALL J=1,N } \\
& \quad A(1: N, J)=A(1: N, J)+B(1: N, K) * C(K, J)
\end{aligned}
$$

## Vectorization with Loop Shifting

if $p_{i}$ is cyclic then
if $k$ is the deepest loop in $\mathrm{P}_{\mathrm{i}}$
then try_recurrence_breaking ( $\left.p_{i}, ~ D, k\right)$
else begin
select_loop_and_interchange ( $\mathrm{p}_{\mathrm{i}}, \mathrm{D}, \mathrm{k}$ ) ; generate a level-k DO statement;
let $D_{i}$ be the dependence graph consisting of
all dependence edges in $D$ that are at level
$k+1$ or greater and are internal to $p_{i}$;
codegen $\left(p_{i}, k+1, D_{i}\right)$;
generate the level-k ENDDO statement end
end

## Scalar Expansion



$$
\square \begin{array}{ll}
S_{1} & T \$(1: N)=A(1: N) \\
S_{2} & A(1: N)=B(1: N) \\
S_{3} & \begin{array}{l}
B(1: N)=T \$(1: N) \\
T=T \$(N)
\end{array}
\end{array}
$$

- Goal: remove anti-dependences inside loops
- Use a different memory location (indexed by loop iterations) for each new value
- Can eliminate dependence cycles inside loops
- Not profitable is scalar variables carry true dependences
- Dependences due to reuse of values must be preserved


## Profitability of Scalar Expansion

- Consider:

$$
\begin{aligned}
& \text { DO } I=1, N \\
& T=T+A(I)+A(I+1) \\
& A(I)=T
\end{aligned}
$$

ENDDO

- Scalar expansion gives us:

$T \$(0)=T$
DO $I=1, N$
$\mathrm{S}_{1}$
$T \$(I)=T \$(I-1)+A(I)+A(I+1)$
$\mathrm{S}_{2}$
$A(I)=T \$(I)$
ENDDO
$T=T \$(N)$
- Cannot eliminate the dependence cycle


## Scalar Expansion: Tradeoffs

- Expansion increases memory requirements
- Solutions:
- Expand in a single loop
- Strip mine loop before expansion
- Forward substitution:

$$
\begin{aligned}
& \mathrm{DO} I=1, N \\
& \mathrm{~T}=\mathrm{A}(\mathrm{I})+\mathrm{A}(\mathrm{I}+1) \\
& \mathrm{A}(\mathrm{I})=\mathrm{T}+\mathrm{B}(\mathrm{I}) \\
& \text { ENDDO } \\
& \text { DO } \mathrm{I}=1, \mathrm{~N} \\
& \mathrm{~A}(\mathrm{I})=\mathrm{A}(\mathrm{I})+\mathrm{A}(\mathrm{I}+1)+\mathrm{B}(\mathrm{I}) \\
& \text { ENDDO }
\end{aligned}
$$

After strip-mining
DO $I 1=1, N, 10$
DO $I=I 1, I 1+9$
$T=A(I)+A(I+1)$
$A(I)=T+B(I)$
ENDDO
ENDDO
DO $\mathrm{I}=\mathrm{I} 1, \mathrm{I} 1+9$
$\mathrm{T}=\mathrm{A}(\mathrm{I})+\mathrm{A}(\mathrm{I}+1)$
$A(I)=T+B(I)$
ENDDO
ENDDO

## Scalar Expansion: Covering Definitions

- A definition $S$ of variable $x$ is a covering definition for loop $L$
- If no other definition of $x$ at the beginning of $L$ can reach uses of $x(S)$ in $L$
- That is, if inside $L$, all uses of $x$ reachable from $S$ has a single definition $S$ (can we apply forward expression substitution?)

DO I $=1,100$
S1 $\quad \mathrm{T}=\mathrm{X}(\mathrm{I})$
S2 $\quad \mathrm{Y}(\mathrm{I})=\mathrm{T}$
ENDDO
DO I = 1, 100
IF (A(I) .GT. 0) THEN
S1 $\quad \mathrm{T}=\mathrm{X}(\mathrm{I})$
S2 $\quad \mathrm{Y}(\mathrm{I})=\mathrm{T}$ ENDIF

$$
Y(I)=T
$$

ENDDO

## Scalar Expansion: Covering Definitions

- A single covering definition may not exist for a loop L
- To form a collection of covering definitions, we can insert dummy assignments:

DO I = 1, 100
IF (A(I) .GT. 0) THEN
S1 $\quad \mathrm{T}=\mathrm{X}(\mathrm{I})$
ELSE
S2 $\quad \mathrm{T}=\mathrm{T}$
ENDIF
S3 $Y(I)=T$
ENDDO

- To compute a set of covering definitions for variable $x$ in $L$
- Find the first definition S1 of $x$ in $L$
- Find all the paths that circumvent S 1 to reach uses of x
- Insert a dummy assignment for $x$ in each of the path found


## Scalar Expansion Using Covering Definitions

$\square$ Given a set C of covering definitions for variable T, assuming loop $L$ has been normalized

- Create an array T\$ of appropriate length
- For each S in the covering definition collection C,
- replace $T$ on the left-hand side by $T \$(\mathrm{I})$.
- For every use of T in the loop body reachable by C
$\square$ If the use is after C in the loop body, replace T by $\mathrm{T} \$(\mathrm{I})$
$\square$ If the use is before C in the loop body, replace T by $\mathrm{T} \$(\mathrm{I}-1)$
- If definitions before the loop $L$ can reach use of $T$ in $L$, insert $\mathrm{T} \$(0)=\mathrm{T}$ before the loop L
- If T is used after loop L , insert $\mathrm{T}=\mathrm{T} \$(\mathrm{U})$ after the loop, where $U$ is the loop upper bound


## Scalar Expansion: Covering Definitions

```
    DO \(I=1,100\)
        IF (A (I) . GT. O) THEN
        \(T=X(I)\)
        ENDIF
\(S_{2} \quad Y(I)=T\)
ENDDO
ENDIF
\(S_{2} \quad Y(I)=T\)
ENDDO
```

$S_{1}$

After inserting covering definitions:
DO $I=1,100$
$S_{1}$
IF (A(I).GT. O) THEN

```
    ELSE
    ENDIF
    \(Y(I)=T\)
```

$\mathrm{S}_{2} \quad \mathrm{~T}=\mathrm{T}$
ENDDO

After scalar expansion:
$T \$(0)=T$
DO $I=1,100$
IF (A (I) . GT. O) THEN
$S_{1} \quad T \$(I)=X(I)$
ELSE
$T \$(I)=T \$(I-1)$
ENDIF
$S_{2} \quad Y(I)=T \$(I)$
ENDDO

## Scalar Renaming

```
DO I = 1, 100
S1 T = A(I) + B(I)
S
S3 T = D(I) - B(I)
S 4 A(I+1) = T * T
    ENDDO
```

$\mathrm{DO} \mathrm{I}=1,100$
$\mathrm{~S}_{1} \mathrm{~T} 1=\mathrm{A}(\mathrm{I})+\mathrm{B}(I)$
$\mathrm{S}_{2} \mathrm{C}(\mathrm{I})=\mathrm{T1}+\mathrm{T1}$
$\mathrm{~S}_{3} \mathrm{~T}=\mathrm{D}(\mathrm{I})-\mathrm{B}(\mathrm{I})$
$\mathrm{S}_{4} \mathrm{~A}(\mathrm{I}+1)=\mathrm{T} 2 * \mathrm{~T} 2$
ENDDO
$\square$

| $S_{3}$ | $T 2 \$(1: 100)=\mathrm{D}(1: 100)-\mathrm{B}(1: 100)$ |
| :--- | :--- |
| $\mathrm{S}_{4}$ | $\mathrm{~A}(2: 101)=\mathrm{T} 2 \$(1: 100) * \mathrm{~T} 2 \$(1: 100)$ |
| $\mathrm{S}_{1}$ | $\mathrm{~T} 1 \$(1: 100)=\mathrm{A}(1: 100)+\mathrm{B}(1: 100)$ |
| $\mathrm{S}_{2}$ | $\mathrm{C}(1: 100)=\mathrm{T} 1 \$(1: 100)+\mathrm{T} \$(1: 100)$ |
|  | $\mathrm{T}=\mathrm{T} 2 \$(100)$ |

- Goal: partition defs/uses into equivalent classes, each of which can occupy different memory locations:
- Pick a definition $S$, add all uses that $S$ reaches
- Add all definitions that reach any of the uses...
- .. until fixed point is reached
- Often done by compilers when calculating live ranges for register allocation


## Array Renaming

DO $I=1, N$
$\mathrm{S}_{1} \quad \mathrm{~A}(\mathrm{I})=\mathrm{A}(\mathrm{I}-1)+\mathrm{X}$
$S_{2} \quad Y(I)=A(I)+Z$
$S_{3} \quad A(I)=B(I)+C$
ENDDO
$-S_{1} \delta_{\infty} S_{2} \quad S_{2} \delta_{\infty}{ }^{-1} S_{3} \quad S_{3} \delta_{1} S_{1} \quad S_{1} \delta_{\infty}{ }^{0} S_{3}$

- Rename A(I) to A\$(I):

DO $I=1, N$

| $S_{1}$ | $A \$(I)=A(I-1)+X$ |
| :--- | ---: |
| $S_{2}$ | $Y(I)=A \$(I)+Z$ |
| $S_{3}$ | $A(I)=B(I)+C$ |

- Dependences remaining: $S_{1} \delta_{\infty} S_{2}$ and $S_{3} \delta_{1} S_{1}$


## Array Renaming: Profitability

$\square$ Examining dependence graph and determining minimum set of critical edges to break a recurrence is NP-complete!
$\square$ Solution:

- Determine edges that are removed by array renaming
- Analyze effects on dependence graph
- Algorithm (assumes no control flow in loop body)
- Identify collections of array references which refer to the same value
- Identify output and anti-dependences to eliminate
- When renaming arrays, minimize amount of copying back to the "original" array at the beginning and the end


## So Far.

- Uncovering potential vectorization in loops by
- Loop Distribution
- Loop Interchange
- Scalar Expansion
- Scalar and Array Renaming
- More transformations
- Loop Skewing
- Node Splitting
- Recognition of Reductions
- Index-Set Splitting
- Run-time Symbolic Resolution
- Putting it together


## Loop Skewing

- Reshape Iteration Space to uncover parallelism
DO I = 1, N
DO J = 1, N
$(=, \leq)$
$S: A(I, J)=A(I-1, J)+A(I, J-1)$ $(<,=) \longrightarrow$
ENDDO
ENDDO
- Dependence Matrix
$\left(\begin{array}{ll}1 & 0 \\ 0 & 1\end{array}\right)$

- Parallelism not apparent


## Loop Skewing Transformation

- Skew iterations of inner loop based on outer loop
- J goes from I+1,I+N instead of 1,N
DO I $=1, \mathrm{~N}$
DO $\mathrm{j}=\mathrm{I}+1, \mathrm{I}+\mathrm{N}$

$$
(=,<)
$$

S: $A(I, j \xrightarrow{(\mathrm{I})=A(\mathrm{I}-1, j-\mathrm{I})}+A(\mathrm{I}, \mathrm{j}-\mathrm{I}-1)$

$$
(\overline{<,<)}
$$

ENDDO
ENDDO

- NOTE: dependence matrix changes

$$
\left(\begin{array}{ll}
1 & 0 \\
0 & 1
\end{array}\right) *\left(\begin{array}{ll}
1 & 1 \\
0 & 1
\end{array}\right)=\left(\begin{array}{ll}
1 & 1 \\
0 & 1
\end{array}\right)
$$

$$
I=3
$$

## Loop Skewing + Loop Interchange

DO I $=1, N$
DO j $=I+1, I+N$
S: $\quad A(I, j-I)=A(I-1, j-I)+A(I, j-I-1)$ ENDDO
ENDDO

## Loop interchange to..

DO $j=2, N+N$
DO I $=\max (1, j-N), \min (N, j-1)$
S: $\quad A(I, j-I)=A(I-1, j-I)+A(I, j-I-1)$ ENDDO
ENDDO

## Vectorize to..

DO j = 2, N+N
FORALL I $=\max (1, j-N), \min (N, j-1)$
S: $\quad A(I, j-I)=A(I-1, j-I)+A(I, j-I-1)$ END FORALL

ENDDO

- Disadvantages:
- After interchange, inner loop evaluates different numbers of iterations
- Outer loop needs twice as much number of iterations
- Not profitable if N is small
- If vector startup time is more than speedup time, this is not profitable
- Vector bounds must be recomputed on each iteration of outer loop
- Apply loop skewing if everything else fails


## Node Splitting



ENDDO

- Recurrence kept intact by renaming algorithm
- Antidependence and true dependence involving the same statement
- Make copy of the source data of antidependence
- Anti-dependence now involves a different stmt
- Goal: break dependence cycle


ENDDO

$$
\begin{aligned}
& X \$(1: N)=X(2: N+1) \\
& X(2: N+1)=B(1: N)+32 \\
& A(1: N)=X \$(1: N)+X(1: N)
\end{aligned}
$$

## Node Splitting

$\square$ Determining minimal set of critical antidependences is in NP-C

- Perfect job of Node Splitting is difficult
- Heuristic:
- Select antidependences
- Delete it to see if acyclic
- If acyclic, apply Node Splitting


## Recognition of Reductions

- Reducing an array of values into a single value
- Sum, min/max, count reductions

$$
S=0.0
$$

$$
\mathrm{DO} \mathrm{I}=1, \mathrm{~N}
$$

ENDDO

$$
\mathrm{S}=\mathrm{S}+\mathrm{A}(\mathrm{I}) \quad \text { Not directly vectorizable }
$$

- Assuming commutativity and associativity

$$
\begin{aligned}
& S=0.0 \\
& \text { DO } k=1,4 \\
& \quad \text { SUM }(k)=0.0 \\
& \text { ENDDO } \\
& \text { DO } I=1, N, 4 \\
& \quad \operatorname{SUM}(1: 3)=\text { SUR } \\
& \text { ENDDO } \\
& \text { DO } k=1,4 \\
& \quad S=S+\text { SUM(k) } \\
& \text { ENDDO }
\end{aligned}
$$

$$
\operatorname{SUM}(1: 3)=\operatorname{SUM}(1: 3)+A(I: I+3)
$$

Useful for vector machines with 4 stage pipeline

## Recognition of Reductions

$\square$ Reduction recognized by

- Presence of self true, output and anti dependences
- Absence of other true dependences

DO I = 1, N


ENDDO

$$
\begin{aligned}
& \text { DO } \mathrm{I}=1, \mathrm{~N} \\
& \mathrm{~S}=\mathrm{S}+\mathrm{A}(\mathrm{I}) \\
& \mathrm{T}(\mathrm{I})=\mathrm{S} \\
& \text { ENDDO }
\end{aligned}
$$

## Index-set Splitting

- Subdivide loop into different iteration ranges to achieve partial parallelization
- Loop Peeling [Weak Zero SIV]
- Threshold Analysis [Strong SIV, Weak Crossing SIV]
- Section Based Splitting [Variation of loop peeling]
- Loop Peeling
- Source of dependence is a single iteration
DO $\mathrm{I}=1, \mathrm{~N}$

$$
A(I)=A(I)+A(1)
$$

ENDDO

```
Loop peeled to..
\(\mathrm{A}(1)=\mathrm{A}(1)+\mathrm{A}(1)\)
DO \(I=2, N\)
    \(A(I)=A(I)+A(1)\)
```

ENDDO
Vectorize to..

$$
\begin{aligned}
& A(1)=A(1)+A(1) \\
& A(2: N)=A(2: N)+A(1)
\end{aligned}
$$

## Threshold Analysis

- Threshold Analysis

DO $\mathrm{I}=1$, 100

$$
A(I+20)=A(I)+B
$$

ENDDO
Strip mine to..
DO I = 1, 100, 20
DO i = I, I+19

$$
A(i+20)=A(i)+B
$$

ENDDO
ENDDO
Vectorize to..
DO I = 1, 100, 20
A(I+20:I+39) =
A(I:I+19)+B

- Crossing thresholds

```
DO I = 1, 100
    A(100-I) = A(I) + B
ENDDO
Strip mine to..
DO I = 1, 100, 50
        DO i = I, I+49
            A(101-i) = A(i) + B
        ENDDO
ENDDO
```

```
Vectorize to..
```

Vectorize to..
DO I = 1, 100, 50
DO I = 1, 100, 50
A(101-I:51-I) = A(I:I+49)+B
A(101-I:51-I) = A(I:I+49)+B
ENDDO

```
ENDDO
```


## Section-based Splitting

```
DO I = 1, N
    DO J = 1, N/2
S1: \(B(J, I)=A(J, I)+C\)
    ENDDO
    DO J = 1, \(N\)
S2: \(A(\mathrm{~J}, \mathrm{I}+1)=\mathrm{B}(\mathrm{J}, \mathrm{I})+\mathrm{D}\)
    ENDDO
ENDDO
- J Loop bound by recurrence due to \(B\)
- Only a portion of B is responsible for it
```

- Partition second loop into loop that uses result of S1 and loop that does not
DO I = 1, N DO J = 1, N/2
S1: $\mathrm{B}(\mathrm{J}, \mathrm{I})=\mathrm{A}(\mathrm{J}, \mathrm{I})+\mathrm{C}$ ENDDO
DO J = 1, N/2
S2: $\mathrm{A}(\mathrm{J}, \mathrm{I}+1)=\mathrm{B}(\mathrm{J}, \mathrm{I})+\mathrm{D}$ ENDDO DO J = N/2+1, N
S3: $A(\mathrm{~J}, \mathrm{I}+1)=\mathrm{B}(\mathrm{J}, \mathrm{I})+\mathrm{D}$ ENDDO
ENDDO


## Run-time Symbolic Resolution

- Breaking conditions

DO $I=1, N$

$$
A(I+L)=A(I)+B(I)
$$

ENDDO
Transformed to..
IF (L.LE.0) THEN
$A(L: N+L)=A(1: N)+B(1: N)$
ELSE
DO $I=1, N$
$A(I+L)=A(I)+B(I)$
ENDDO
ENDIF

- Identifying minimum number of breaking conditions to break a recurrence is in NPComplete
- Heuristic:
- Identify when a critical dependence can be conditionally eliminated via a breaking condition


## Putting It All Together

- Good Part
- Many transformations imply more choices to exploit parallelism
- Bad Part
- Choosing the right transformation
- How to automate transformation selection?
- Interference between transformations
- An effective optimization algorithm must
- Take a global view of transformed code
- Know the architecture of the target machine
- Example of Interference

```
DO I = 1, N
    DO J = 1, M
        S(I) = S(I) + A(I,J)
    ENDDO
```

ENDDO

## Sum Reduction gives..

DO $\mathrm{I}=1$, N $S(I)=S(I)+S U M(A(I, 1: M))$
ENDDO
While Loop Interchange and Vectorization gives..
DO $\mathrm{J}=1$, N
$S(1: N)=S(1: N)+A(1: N, J)$
ENDDO

## Performance on Benchmark

| Vectorizing Compiler | Total |  |  | Dependence |  |  | Vectorization |  |  | Idioms |  |  | Completeness |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | V | P | N | V | P | N | V | P | N | V | P | N | V | P | N |
| PFC | 70 | 6 | 24 | 17 | 0 | 7 | 25 | 4 | 5 | 5 | 0 | 10 | 23 | 2 | 2 |
| Alliant FX/8, Fortran V4.0 | 68 | 5 | 27 | 19 | 0 | 5 | 20 | 5 | 9 | 10 | 0 | 5 | 19 | 0 | 8 |
| Amdahl VP-E, Fortran 77 | 62 | 11 | 27 | 16 | 1 | 7 | 21 | 8 | 5 | 11 | 1 | 3 | 14 | 1 | 12 |
| Ardent Titan-1 | 62 | 6 | 32 | 18 | 0 | 6 | 19 | 5 | 10 | 9 | 0 | 6 | 16 | 1 | 10 |
| CDC Cyber 205, VAST-2 | 62 | 5 | 33 | 16 | 0 | 8 | 20 | 5 | 9 | 7 | 0 | 8 | 19 | 0 | 8 |
| CDC Cyber 990E/995E | 25 | 11 | 64 | 8 | 0 | 16 | 6 | 8 | 20 | 3 | 1 | 11 | 8 | 2 | 17 |
| Convex C Series, FC 5.0 | 69 | 5 | 26 | 17 | 0 | 7 | 25 | 4 | 5 | 11 | 0 | 4 | 16 | 1 | 10 |
| Cray series, CF77 V3.0 | 69 | 3 | 28 | 20 | 0 | 4 | 18 | 3 | 13 | 9 | 0 | 6 | 22 | 0 | 5 |
| CRAX X-MP , CFT V1.15 | 50 | 1 | 49 | 16 | 0 | 8 | 12 | 1 | 21 | 10 | 0 | 5 | 12 | 0 | 15 |
| Cray Series, CFT77 V3.0 | 50 | 1 | 49 | 17 | 0 | 7 | 8 | 1 | 25 | 7 | 0 | 8 | 18 | 0 | 9 |
| CRAY-2, CFT2 V3.1a | 27 | 1 | 72 | 5 | 0 | 19 | 3 | 1 | 30 | 8 | 0 | 7 | 11 | 0 | 16 |
| ETA-10, FTN 77 V1.0 | 62 | 7 | 31 | 18 | 0 | 6 | 18 | 7 | 9 | 7 | 0 | 8 | 19 | 0 | 8 |
| Gould NP1, GCF 2.0 | 60 | 7 | 33 | 14 | 0 | 10 | 19 | 7 | 8 | 8 | 0 | 7 | 19 | 0 | 8 |
| Hitachi S-810/820, | 67 | 4 | 29 | 14 | 0 | 10 | 24 | 4 | 6 | 14 | 0 | 1 | 15 | 0 | 12 |
| IBM 3090/VF, VS Fortran | 52 | 4 | 44 | 12 | 0 | 12 | 19 | 3 | 12 | 5 | 1 | 9 | 16 | 0 | 11 |
| Intel iPSC/2-VX, VAST-2 | 56 | 8 | 36 | 15 | 0 | 9 | 17 | 8 | 9 | 6 | 0 | 9 | 18 | 0 | 9 |
| NEC SX/2, F77/SX | 66 | 5 | 29 | 17 | 0 | 7 | 21 | 5 | 8 | 12 | 0 | 3 | 16 | 0 | 11 |
| SCS-40, CFT x13g | 24 | 1 | 75 | 7 | 0 | 17 | 6 | 1 | 27 | 5 | 0 | 10 | 6 | 0 | 21 |
| Stellar GS 1000, F77 | 48 | 11 | 41 | 14 | 0 | 10 | 20 | 9 | 5 | 4 | 1 | 10 | 10 | 1 | 16 |
| Unisys ISP, UFTN 4.1.2 | 67 | 13 | 20 | 21 | 3 | 0 | 19 | 8 | 7 | 10 | 2 | 3 | 17 | 0 | 10 |

