Getting the Most Out of the FFTs in the Cray X1 Scientific Library

Bracy Elton, Ph.D.
May 16, 2003
I. FFT/Convolutions/Filtering Library Overview.
II. Porting Issues.
III. Performance Issues.
IV. Performance/Timing Measurements.
V. Future Plans.
VI. Conclusions.
• Address porting from previous Cray & other systems, e.g., Cray T90, Cray SV1, Cray T3E, and workstation-type systems.

• Support more data types.
(I) FFT Library Overview: Variants

- Default LibSci.
  - default Scientific libraries variant.
  - 32-bit integers.
  - `-s default32` (the default) or `-lsci` or `-lsci32`.
  - Single & double precision names.

- LibSci (64-bit).
  - Scientific libraries variant most compatible with previous Cray systems.
  - 64-bit integers.
  - `-s default64` or `-lsci64`.
  - Single precision names only.

- Single MSP & single SSP versions.
## FFT Library Overview: Data Types

<table>
<thead>
<tr>
<th>Library</th>
<th>Integer Width</th>
<th>Floating Point Precision</th>
<th>Floating Point Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>LibSci (default)</td>
<td>32 bits</td>
<td>Single</td>
<td>32 bits</td>
</tr>
<tr>
<td>LibSci (default)</td>
<td>32 bits</td>
<td>Double</td>
<td>64 bits</td>
</tr>
<tr>
<td>LibSci (64-bit)</td>
<td>64 bits</td>
<td>Single</td>
<td>64 bits</td>
</tr>
</tbody>
</table>
(I) FFT Library Overview: Documentation

• Man pages:
  – intro_libsci.
  – intro_fft.

• Manuals:
  – *Cray X1 User Environment Differences*.
  – *Migrating Applications to Cray X1 Systems*. 
More manuals coming:

- *Optimizing Applications on the Cray X1 System* to have additional chapter for using LibSci.
(I) FFT Library Overview: Contents

• 1-D, 2-D, 3-D, multiple 1-D complex-to-complex, real-to-complex and complex-to-real FFTs/DFTs.

• Convolutions.
  – Directly computed.
  – Computed via FFTs.

• Filters.
  – Correlation of two vectors with general coefficient.
  – Correlation of two vectors with symmetric coefficient.
  – Weiner-Levinson linear equations solution.
## Single Precision FFTs

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Complex-to-complex</th>
<th>Real-to-Complex</th>
<th>Complex-to-Real</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-D</td>
<td>CCFFT (CFFT)</td>
<td>SCFFT</td>
<td>CSFFT</td>
</tr>
<tr>
<td>2-D</td>
<td>CCFFT2D (CFFT2D)</td>
<td>SCFFT2D</td>
<td>CSFFT2D</td>
</tr>
<tr>
<td>3-D</td>
<td>CCFFT3D (CFFT3D)</td>
<td>SCFFT3D</td>
<td>CSFFT3D</td>
</tr>
<tr>
<td>Multiple 1-D</td>
<td>CCFFTM (MCFFT)</td>
<td>SCFFTM</td>
<td>CSFFTM</td>
</tr>
</tbody>
</table>
## Double Precision FFTs

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Complex-to-complex</th>
<th>Real-to-Complex</th>
<th>Complex-to-Real</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-D</td>
<td>ZZFFT (ZFFT)</td>
<td>DZFFT</td>
<td>ZDFFFT</td>
</tr>
<tr>
<td>2-D</td>
<td>ZZFFT2D (ZFFT2D)</td>
<td>DZFFT2D</td>
<td>ZDFFFT2D</td>
</tr>
<tr>
<td>3-D</td>
<td>ZZFFT3D (ZFFT3D)</td>
<td>DZFFT3D</td>
<td>ZDFFFT3D</td>
</tr>
<tr>
<td>Multiple 1-D</td>
<td>ZZFFTM (MZFFT)</td>
<td>DZFFTM</td>
<td>ZDFFTM</td>
</tr>
<tr>
<td>Description</td>
<td>Single Precision</td>
<td>Double Precision</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>Complex Convolution (FFT)</td>
<td>CCNVLF</td>
<td>ZCNVLF</td>
<td></td>
</tr>
<tr>
<td>Direct Complex Convolution</td>
<td>CCNVL</td>
<td>ZCNVL</td>
<td></td>
</tr>
<tr>
<td>Symmetric Correlation</td>
<td>SFILTERS (FILTERS)</td>
<td>DFILTERS</td>
<td></td>
</tr>
<tr>
<td>General Correlation</td>
<td>SFILTERG (FILTERG)</td>
<td>DFILTERG</td>
<td></td>
</tr>
<tr>
<td>Weiner-Levinson Solver</td>
<td>SOPFILT (OPFILT)</td>
<td>DOPFILT</td>
<td></td>
</tr>
</tbody>
</table>
(II) Porting Issues

- **Accuracy.**
  - 32- vs. 64-bit floating point format.

- **FFT & convolution TABLE & WORK space.**
  - Sizes may differ from previous systems.
  - TABLE always array of 64-bit words.
    - Size varies depending on floating point word length.
  - WORK may vary depending on MSP vs. SSP mode & on library and floating point precision.

- **Sign of exponent in Nth root of unity.**
  - Same as on previous Cray systems.
  - May differ on non-Cray systems.
  - ISIGN & SCALE parameter choices span the mathematical possibilities.
(II) Porting Issues (cont.)

• Routine names.
  – Default LibSci vs. LibSci (64-bit).
  – Single vs. double precision.
• MSP vs. SSP mode.
• LibSci mixed radix FFTs:
  – Complex-to-complex radix 2, 3, 4, 5 & 8 butterflies.
  – Real-to-complex/complex-to-real radix 2, 3, 4, 5, 6 & 8 butterflies.
• Compilation flags.
• Linking to desired library.
(III) Performance Issues for FFTs

- FFT Length.
- Strides & leading dimensions.
- Algorithm choice for 3-D FFTs.
- Greatest tuning effort on complex-to-complex FFTs.
FFT Lengths

- Lengths containing factors that are not powers of 2, 3, or 5 result in DFT implementations.
- Powers of 2 generally better than powers of 3 and 5.
- Separate radix 4 and radix 8 butterflies for complex-to-complex transforms.
- Multistreaming increases $N^{1/2}$.
  - ($N^{1/2}$ is length to reach $1/2$ of algorithmic peak.)
  - Longer vectors better.
- Check nearby sizes, if situation allows.
Strides and Leading Dimensions

- Use stride information to change leading dimensions.
- Consider 128**3 case.
- Power of 2 strides bad.

\[
\text{COMPLEX*16 } X(128, 128, 128)
\]

- Odd multiples of 4 (8 for 32-bit floating point data) better.

\[
\text{COMPLEX*16 } X(130, 129, 128)
\]

- Odd strides better (odd multiple of 2 leading dimension okay).

\[
\text{COMPLEX*16 } X(129, 129, 128)
\]
Leading Dimensions/Stride Examples

- LibSci (64-bit).
- Forward complex-to-complex.
- 2-D, 3-D, Multiple 1-D.
- Power of 2 sizes.
- No leading dimension adjustments vs. using odd leading dimensions.
  - All but last dimension changed to be odd.
Time vs. Performance

- Performance in evaluation of FFTs:
  - Radices used influence operation count.
  - Normalize operation count via theoretical value.
  - Complex-to-complex Cooley-Tukey operation count:

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Operation Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-D</td>
<td>$5<em>N</em>\log_2(N)$</td>
</tr>
<tr>
<td>2-D</td>
<td>$5<em>N_1</em>N_2*\log_2(N_1*N_2)$</td>
</tr>
<tr>
<td>3-D</td>
<td>$5<em>N_1</em>N_2<em>N_3</em>\log_2(N_1<em>N_2</em>N_3)$</td>
</tr>
<tr>
<td>Multiple 1-D</td>
<td>$5<em>N</em>M*\log_2(N)$</td>
</tr>
</tbody>
</table>

- Real-to-complex/complex-to-real Cooley-Tukey count:
  - Substitute $N/2$ for $N$ and $(N_1)/2$ for $N_1$ in above.

- Time is what really counts for FFTs.
CCFFT3D (64-bit, MSP) Timings

LibSci CCFFT3D (64-bit, MSP, ISYS=0)

MSP, ISYS=0, No adjustment

MSP, ISYS=0, Odd leading dimensions

Time (sec.)

N (=N1=N2=N3)

2 4 8 16 32 64 128 256 512

0.00001

0.0001

0.001

0.01

0.1

1

10

100
CCFFFT3D (64-bit, MSP) Timings

LibSci CCFFFT3D (64-bit, MSP, ISYS=0)

- MSP, ISYS=0, No adjustment
- MSP, ISYS=0, Odd leading dimensions

Time (sec.) vs. N (=N1=N2=N3)

2 4 8 16 32 64 128 256 512
CCFFT3D (64-bit, MSP) Performance

LibSci CCFFT3D (64-bit, MSP, ISYS=0)

MSP, ISYS=0, No adjustment
MSP, ISYS=0, Odd leading dimensions

Performance (Cooley-Tukey Mflop/s)

N (N1=N2=N3)

2 4 8 16 32 64 128 256 512
CCFFT3D (64-bit, MSP) Performance

LibSci CCFFT3D (64-bit, MSP, ISYS=0)

MSP, ISYS=0, No adjustment
MSP, ISYS=0, Odd leading dimensions

Performance (Cooley-Tukey Mflop/s)

N (=N1=N2=N3)

0 500 1000 1500 2000 2500 3000 3500 4000 4500

2 4 8 16 32 64 128 256 512
CCFFT2D (64-bit, MSP) Timings

LibSci CCFFT2D (64-bit, MSP, ISYS=0)

MSP, ISYS=0, No adjustment
MSP, ISYS=0, Odd leading dimensions
CCFFT2D (64-bit, MSP) Timings

LibSci CCFFT2D (64-bit, MSP, ISYS=0)

- **MSP, ISYS=0, No adjustment**
- **MSP, ISYS=0, Odd leading dimensions**

Time (sec.) vs. \( N \) (=\( N1=N2 \))

- **N** = 1, 4, 16, 64, 256, 1024, 4096

Graph showing performance over varying sizes of \( N \).
CCFFT2D (64-bit, MSP) Performance

LibSci CCFFT2D (64-bit, MSP, ISYS=0)

Performance (Cooley-Tukey Mflop/s)

N (=N1=N2)

MSP, ISYS=0, No adjustment
MSP, ISYS=0, Odd leading dimensions
CCFFT2D (64-bit, MSP) Performance

LibSci CCFFT2D (64-bit, MSP, ISYS=0)

Performance (Cooley-Tukey Mflop/s)

N (=N1=N2)

MSP, ISYS=0, No adjustment
MSP, ISYS=0, Odd leading dimensions
CCFFTM (64-bit, MSP) Timings

LibSci CCFFTM (64-bit, MSP, ISYS=0)

MSP, ISYS=0, No adjustment
MSP, ISYS=0, Odd leading dimensions

Time (sec.)

N (=M)

1e-05 0.0001 0.001 0.01 0.1 1 10

1 4 16 64 256 1024 4096

SLIDE 27
Getting the Most Out of the Cray X1 FFTs / Bracy Elton
5/16/2003
CUG 2003 / Columbus, Ohio, USA
CCFFTM (64-bit, MSP) Timings

LibSci CCFFTM (64-bit, MSP, ISYS=0)

- MSP, ISYS=0, No adjustment
- MSP, ISYS=0, Odd leading dimensions

Time (sec.)

N (=M)

1 4 16 64 256 1024 4096
CCFFTM (64-bit, MSP) Performance

LibSci CCFFTM (64-bit, MSP, ISYS=0)

- MSP, ISYS=0, No adjustment
- MSP, ISYS=0, Odd leading dimensions

Performance (Cooley-Tukey Mflop/s) vs N (=M)
CCFFTM (64-bit, MSP) Performance

LibSci CCFFTM (64-bit, MSP, ISYS=0)

Performance (Cooley-Tukey Mflop/s)

- MSP, ISYS=0, No adjustment
- MSP, ISYS=0, Odd leading dimensions

N (=M)

0 500 1000 1500 2000 2500 3000 3500 4000 4500

1 4 16 64 256 1024 4096
Algorithm Choice for 3-D FFTs

- **ISYS=0.**
  - In each of the three 1-D FFTs:
    - 1 dimension for managing less memory.
    - 1 dimension for vectorization & multistreaming.
    - 1 dimension for transform.
- Requires less work space than ISYS=1.
- Generally less performance.
• ISYS=1.
  – In each of the three 1-D FFTs:
    • 1 dimension for multistreaming.
    • 1 dimension for vectorization.
    • 1 dimension for transform.
• Requires more WORK space than ISYS=0.
• Generally more performance.
Algorithm Comparisons

- MSP mode.
- LibSci (64-bit).
- Complex-to-complex routine CCFFT3D.
- N = N1 = N2 = N3 as plotted contains only factors that are powers of 2, 3, and 5.
- Graphs drawn in “continuous” fashion for visual clarity.
CCFFT3D (64-bit, MSP) Timings

LibSci CCFFT3D (64-bit, MSP, Odd leading dimensions)

- ISYS=0, MSP, Odd leading dimensions
- ISYS=1, MSP, Odd leading dimensions
CCFFT3D (64-bit, MSP) Performance

LibSci CCFFT3D (64-bit, MSP)

- ISYS=0, MSP, Odd leading dimensions
- ISYS=1, MSP, Odd leading dimensions

Performance (Cooley-Tukey Mflop/s)

N (=N1=N2=N3)

0 1000 2000 3000 4000 5000 6000

2 4 8 16 32 64 128 256 512
(IV) Performance & Timings

- 1-D FFTs.
- 2-D FFTs.
- 3-D FFTs.
- Multiple 1-D FFTs.
- Complex-to-complex vs. real-to-complex/complex-to-real.
- MSP vs. SSP vs. SV1ex vs. T94.
- $\text{ISYS} = 0$ vs. $\text{ISYS} = 1$.
- No lengths containing factors not powers of 2, 3, and 5.
CCFFT3D (64-bit) Timings

Time (sec.) vs. N (=N1=N2=N3)

- Cray X1 MSP, ISYS=1, Odd leading dimensions
- Cray X1 SSP, ISYS=1, Odd leading dimensions
- Cray SV1ex
CCFFT3D (64-bit) Timings

- Cray X1 MSP, ISYS=1, Odd leading dimensions
- Cray X1 SSP, ISYS=1, Odd leading dimensions
- Cray SV1ex

Time (sec.) vs. N (=N1=N2=N3)

N: 2, 4, 8, 16, 32, 64, 128, 256, 512
CCFFT3D (64-bit) Performance

Cray X1 MSP, ISYS=1, Odd leading dimensions
Cray X1 SSP, ISYS=1, Odd leading dimensions
Cray SV1ex

N (N1=N2=N3)

Performance (Cooley-Tukey Mflop/s)
CCFFT3D (64-bit) Performance

Performance (Cooley-Tukey Mflop/s) vs. N (N1=N2=N3)

- Cray X1 MSP, ISYS=1, Odd leading dimensions
- Cray X1 SSP, ISYS=1, Odd leading dimensions
- Cray SV1ex
CCFFTM (64-bit) Timings

CCFFTM (64-bit)

- Cray X1 MSP, Odd leading dimensions
- Cray X1 SSP, Odd leading dimensions

Time (sec.) vs. N (=M)
**CCFFTM (64-bit) Timings**

![Graph showing CCFFTM (64-bit) Timings]

- **Cray X1 MSP, Odd leading dimensions**
- **Cray X1 SSP, Odd leading dimensions**

**Axes:**
- **X-axis:** N (=M)
- **Y-axis:** Time (sec.)

**Legend:**
- Red line: Cray X1 MSP, Odd leading dimensions
- Green line: Cray X1 SSP, Odd leading dimensions

**Graph Details:**
- The graph plots time (in seconds) against the number of elements (N) for different dimensions of data.
- The curve for Cray X1 MSP, Odd leading dimensions shows a higher time as the data size increases compared to Cray X1 SSP, Odd leading dimensions.
CCFFTM (64-bit) Performance

Performance (Cooley-Tukey Mflop/s) vs. N (=M)

- Cray X1 MSP, Odd leading dimensions
- Cray X1 SSP, Odd leading dimensions
CCFFTM (64-bit) Performance

Performance (Cooley-Tukey Mflop/s) vs. N (=M)

- Cray X1 MSP, Odd leading dimensions
- Cray X1 SSP, Odd leading dimensions
SC/CSFFTM (64-bit) Timings

SCFFTM/CSFFTM (64-bit)

Cray X1 MSP, Odd leading dimensions
Cray X1 SSP, Odd leading dimensions

Time (sec.)

N (=M)

1e-05
0.0001
0.001
0.01
0.1
1
4
16
64
256
1024
4096
SC/CSFFTM (64-bit) Timings

- Red line: Cray X1 MSP, Odd leading dimensions
- Green line: Cray X1 SSP, Odd leading dimensions

Time (sec.) vs. N (=M) for SCFFTM/CSFFTM (64-bit) timings.
SC/CSFFTM (64-bit) Performance

Performance (Cooley-Tukey Mflop/s)

N (=M)

1 4 16 64 256 1024 4096

Cray X1 MSP, Odd leading dimensions
Cray X1 SSP, Odd leading dimensions
SC/CSFFTM (64-bit) Performance

Performance (Cooley-Tukey Mflop/s) vs. N (=M)

- Cray X1 MSP, Odd leading dimensions
- Cray X1 SSP, Odd leading dimensions
CCFFT2D (64-bit) Timings

CCFFT2D (64-bit)

- Cray X1 MSP, Odd leading dimensions
- Cray X1 SSP, Odd leading dimensions
- Cray T94
- Cray SV1ex

Time (sec.) vs. N (=N1=N2)

N values: 1, 4, 16, 64, 256, 1024, 4096, 16384
CCFFT2D (64-bit) Performance

Performance (Cooley-Tukey Mflop/s)

N (=N1=N2)

Cray X1 MSP, Odd leading dimensions
Cray X1 SSP, Odd leading dimensions
Cray T94
Cray SV1ex

0.1 1 10 100 1000 10000
1 4 16 64 256 1024 4096 16384

SLIDE 53
Getting the Most Out of the Cray X1 FFTs / Bracy Elton
5/16/2003
CUG 2003 / Columbus, Ohio, USA
CCFFT2D (64-bit) Performance

Performance (Cooley-Tukey Mflop/s) vs. N (N1 = N2)

- Cray X1 MSP, Odd leading dimensions
- Cray X1 SSP, Odd leading dimensions
- Cray T94
- Cray SV1ex
**CCFFT (64-bit) Timings**

![Graph showing CCFFT (64-bit) Timings](image)

- **N** (Input Size):
  - 1
  - 4
  - 16
  - 64
  - 256
  - $2^{10}$
  - $2^{12}$
  - $2^{14}$
  - $2^{16}$
  - $2^{18}$
  - $2^{20}$

- **Time (sec.)**:
  - $10^{-5}$
  - 0.0001
  - 0.001
  - 0.01
  - 0.1
  - 1

**Curves**:
- **Cray X1 MSP**
- **Cray X1 SSP**
- **Cray T94**
- **Cray SV1ex**

---

**Slide Information**
- **Slide**: 55
- **Title**: Getting the Most Out of the Cray X1 FFTs / Bracy Elton
- **Date**: 5/16/2003
- **Conference**: CUG 2003 / Columbus, Ohio, USA
CCFFT (64-bit) Performance

Performance (Cooley-Tukey Mflop/s) vs. N

- Cray X1 MSP
- Cray X1 SSP
- Cray T94
- Cray SV1ex
CCFFT (64-bit) Performance

Performance (Cooley-Tukey Mflop/s)

N

Cray X1 MSP
Cray X1 SSP
Cray T94
Cray SV1ex
Future Plans

- Further optimization.
  - Possible new algorithms.
  - Better use of cache.
  - Finding sweet spot for various tuning parameters.
  - Better instruction scheduling for complex-to-complex radix 3 and radix 5 butterflies.
  - Perhaps include allocating vs. non-allocating vector loads & stores choice at runtime.
- Fortran90 module interface block for LibSci.
- 2-D & 3-D distributed memory parallel FFTs.
(VI) Conclusions

• Mind porting issues:
  – LibSci variants.
  – Data types & accuracy.
  – TABLE and WORK storage differences.

• Choose FFT lengths wisely.
• Mind ISYS=1 possibilities (3-D currently).
• Increase problem dimensions for more performance.
• Adjust leading dimensions for good strides.