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

Bracy Elton, Ph.D.

May 16, 2003

Outline

I. FFT/Convolutions/Filtering Library Overview.
II. Porting Issues.
III. Performance Issues.
IV. Performance/Timing Measurements.
V. Future Plans.
VI. Conclusions.

(I) FFT Library Overview: Motivation

The 45th CUG Conference
CRAY USER GROUP
FLIGHT TO INSIGHT
May 12-16, 2003 ○ Columbus, Ohio

- 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.**

(I) FFT Library Overview: Data Types

Library	Integer Width	Floating Point Precision	Floating Point Length
LibSci (default)	32 bits	Single	32 bits
LibSci (default)	32 bits	Double	64 bits
LibSci (64-bit)	64 bits	Single	64 bits

(I) FFT Library Overview: Documentation

- **Man pages:**
  - **intro_libsci.**
  - **intro_fft.**
- **Manuals:**
  - ***Cray X1 User Environment Differences.***
  - ***Migrating Applications to Cray X1 Systems.***

(I) FFT Library Overview: Documentation

- **More manuals coming:**
  - ***Optimizing Applications on the Cray X1 System*** **to have additional chapter for using LibSci.**
  - **Cray X1 LibSci reference manual.**

(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

Dimension	Complex-to-complex	Real-to-Complex	Complex-to-Real
1-D	CCFFT (CFFT)	SCFFT	CSFFT
2-D	CCFFT2D (CFFT2D)	SCFFT2D	CSFFT2D
3-D	CCFFT3D (CFFT3D)	SCFFT3D	CSFFT3D
Multiple 1-D	CCFFTM (MCFFT)	SCFFTM	CSFFTM

Double Precision FFTs

Dimension	Complex-to-complex	Real-to-Complex	Complex-to-Real
1-D	ZZFFT (ZFFT)	DZFFT	ZDFFT
2-D	ZZFFT2D (ZFFT2D)	DZFFT2D	ZDFFT2D
3-D	ZZFFT3D (ZFFT3D)	DZFFT3D	ZDFFT3D
Multiple 1-D	ZZFFTM (MZFFT)	DZFFTM	ZDFFTM

Convolutions & Filters

The 45th CUG Conference
CRAY USER GROUP
FLIGHT TO INSIGHT
OSC
May 12-16, 2003 ∘ Columbus, Ohio

Description	Single Precision	Double Precision
Complex Convolution (FFT)	CCNVLF	ZCNVLF
Direct Complex Convolution	CCNVL	ZCNVL
Symmetric Correlation	SFILTERS (FILTERS)	DFILTERS
General Correlation	SFILTERG (FILTERG)	DFILTERG
Weiner-Levinson Solver	SOPFILT (OPFILT)	DOPFILT

(II) Porting Issues

The 45th CUG Conference
CRAY USER GROUP
FLIGHT TO INSIGHT
OSC
May 12-16, 2003 ○ Columbus, Ohio

- 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 $N_{½}$.
  - ($N_{½}$ is length to reach ½ 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.

```
COMPLEX*16 X(128,128,128)
```

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

```
COMPLEX*16 X(130,129,128)
```

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

```
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.

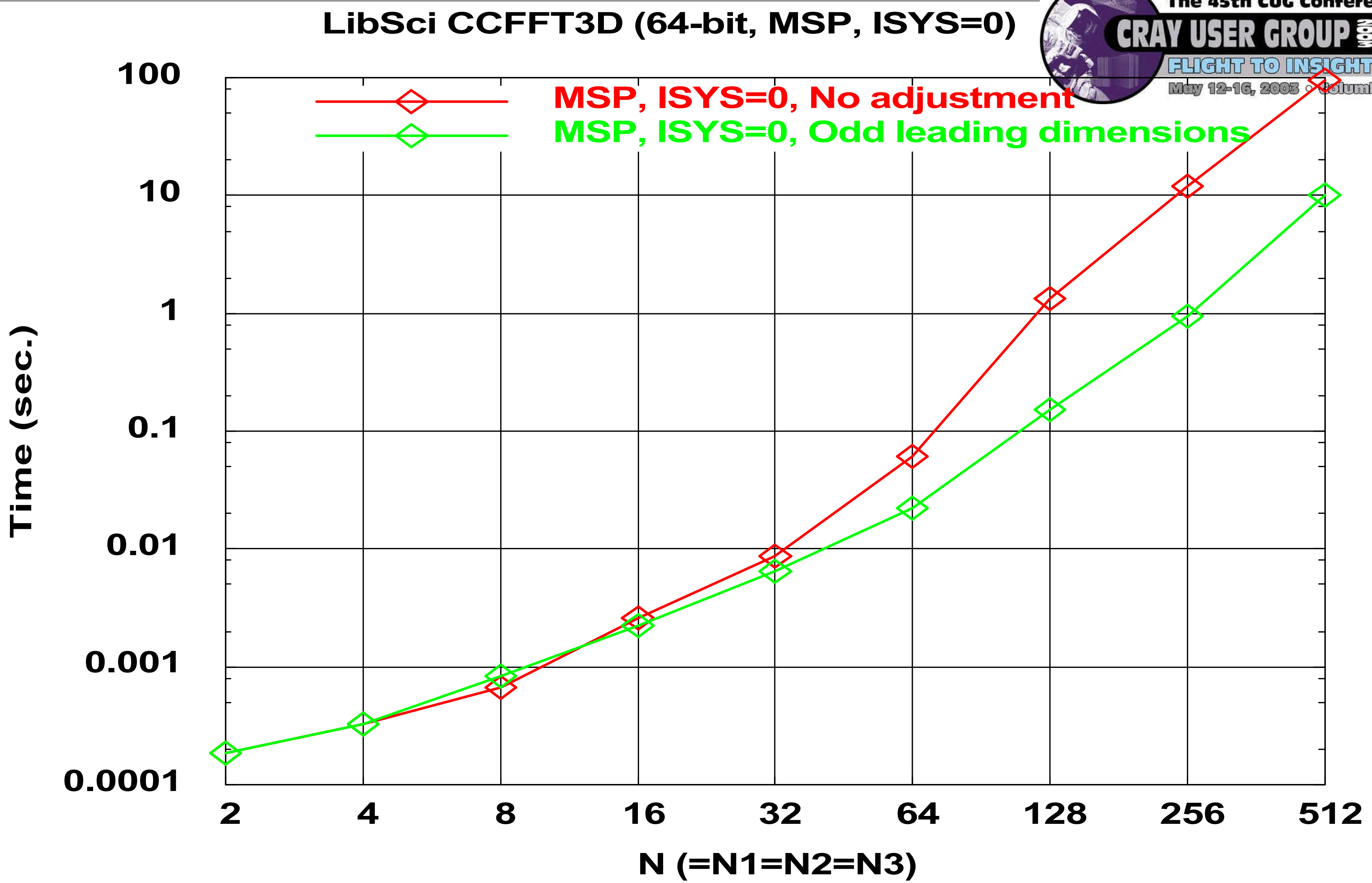
Time vs. Performance

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

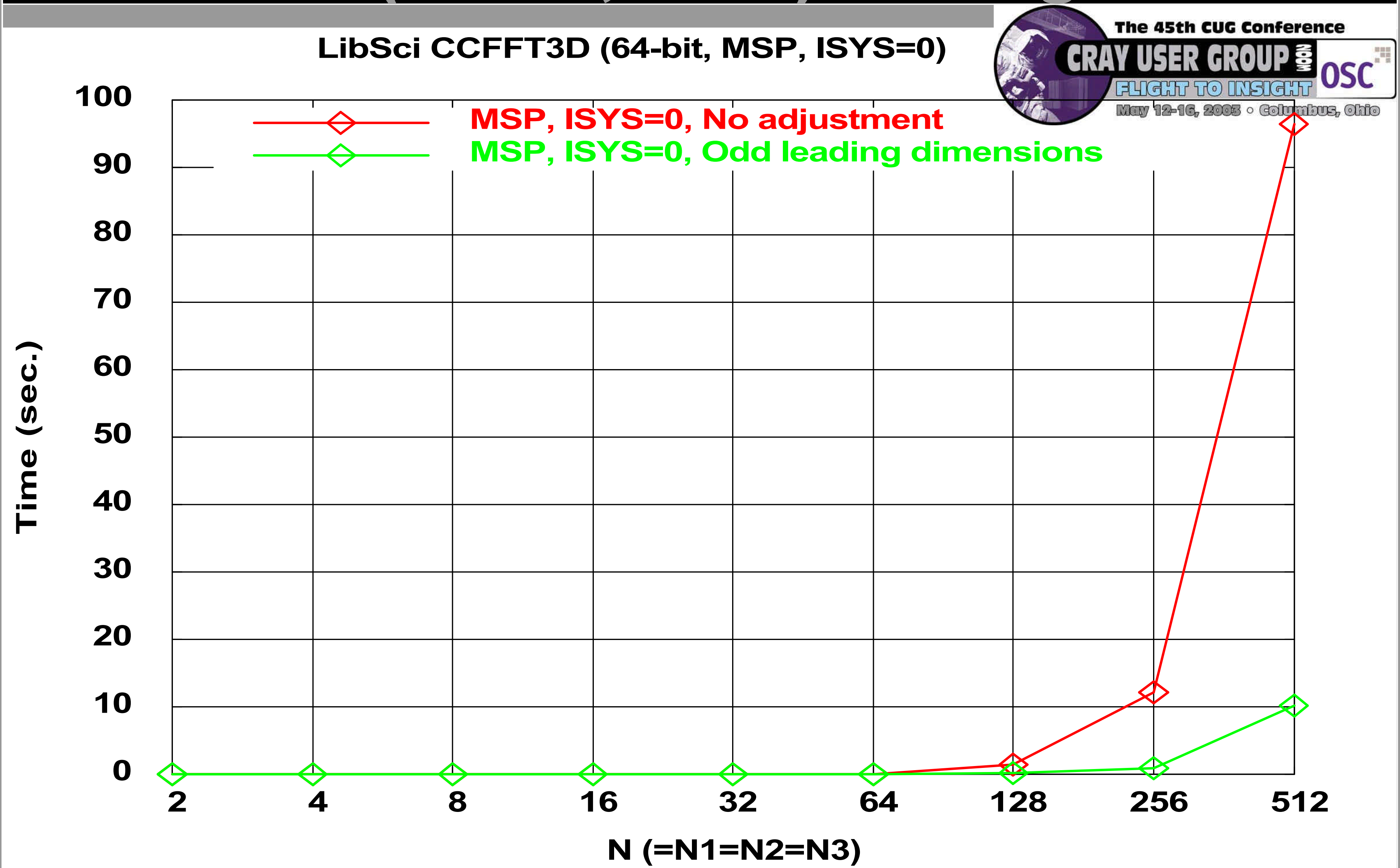
1-D	5*N*log_2(N)
2-D	5*N1*N2*log_2(N1*N2)
3-D	5*N1*N2*N3*log_2(N1*N2*N3)
Multiple 1-D	5*N*M*log_2(N)

  - Real-to-complex/complex-to-real Cooley-Tukey count:
    - Substitute N/2 for N and (N1)/2 for N1 in above.
- Time is what really counts for FFTs.

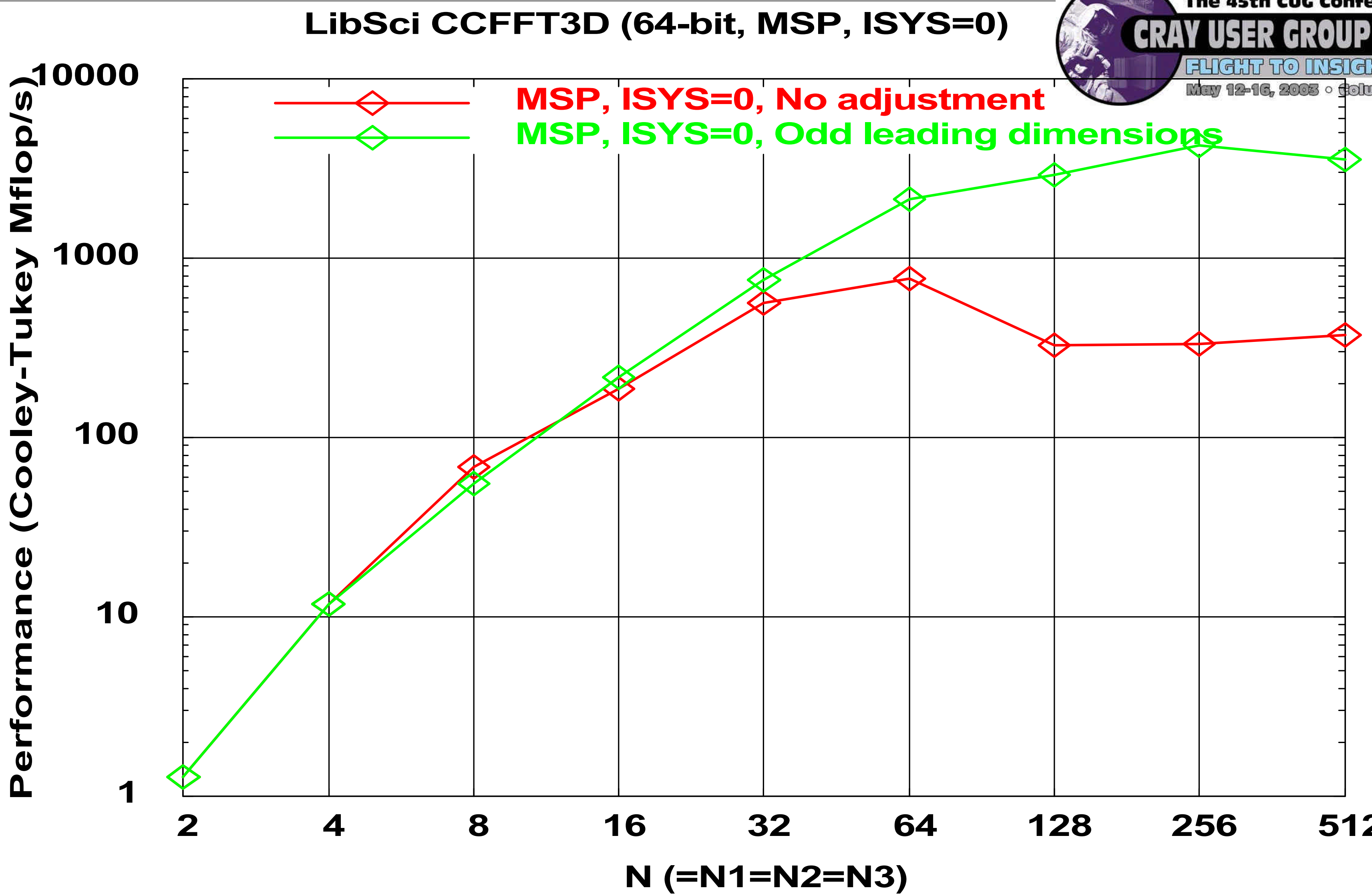
CCFFT3D (64-bit, MSP) Timings



CCFFT3D (64-bit, MSP) Timings

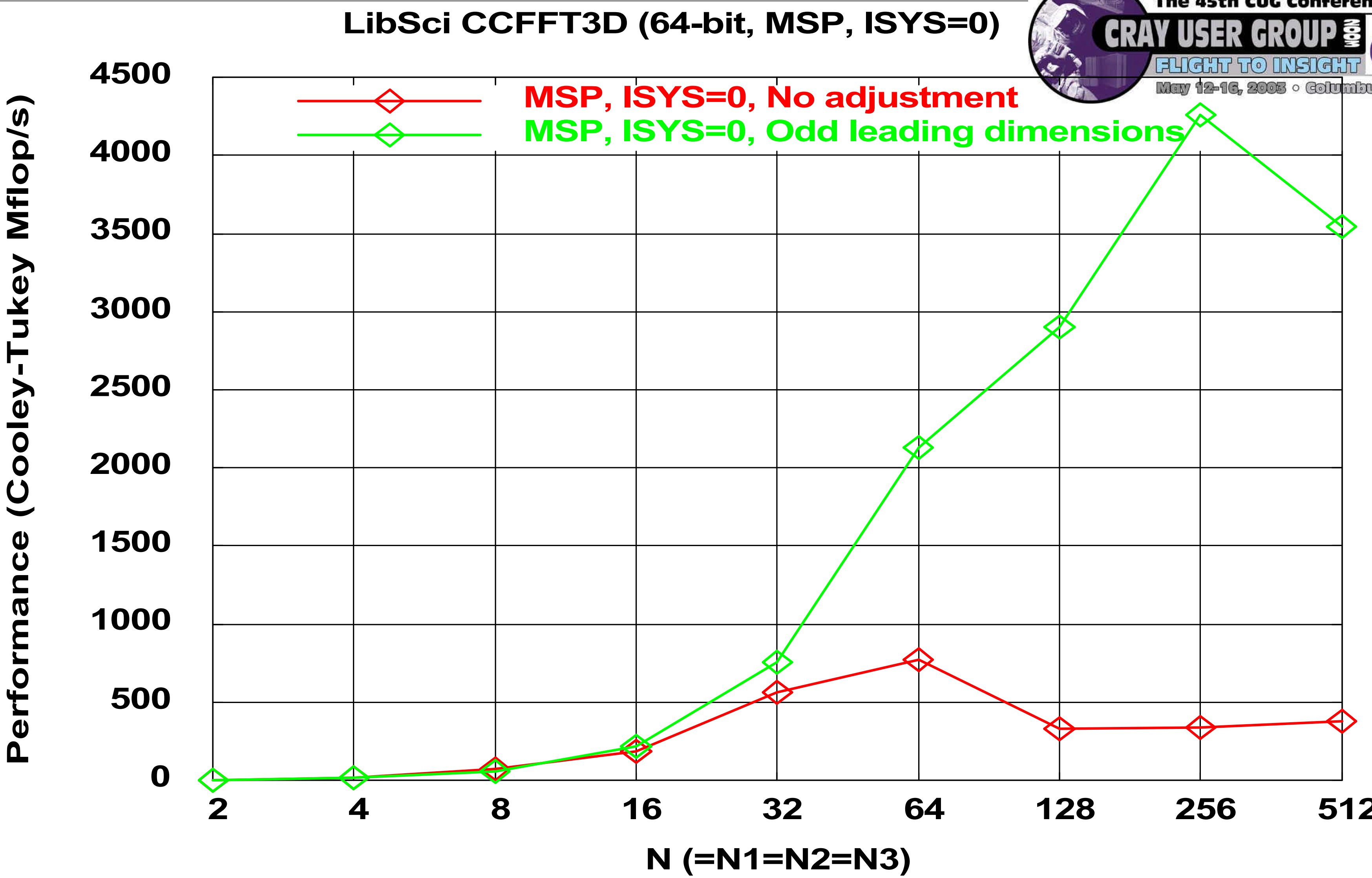


CCFFT3D (64-bit, MSP) Performance

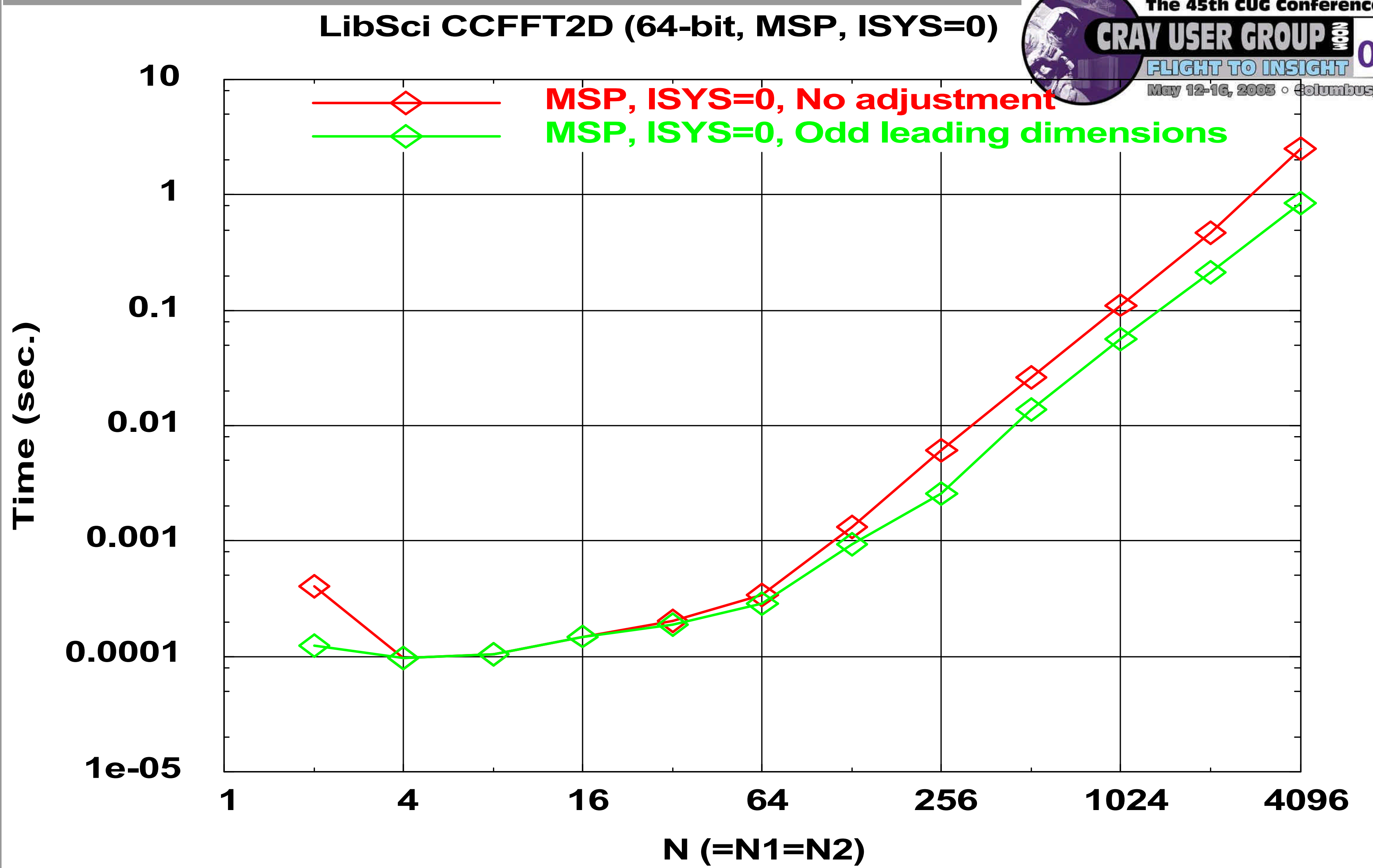


The 45th CUG Conference
CRAY USER GROUP
OSC
FLIGHT TO INSIGHT
May 12-16, 2003 ○ Columbus, Ohio

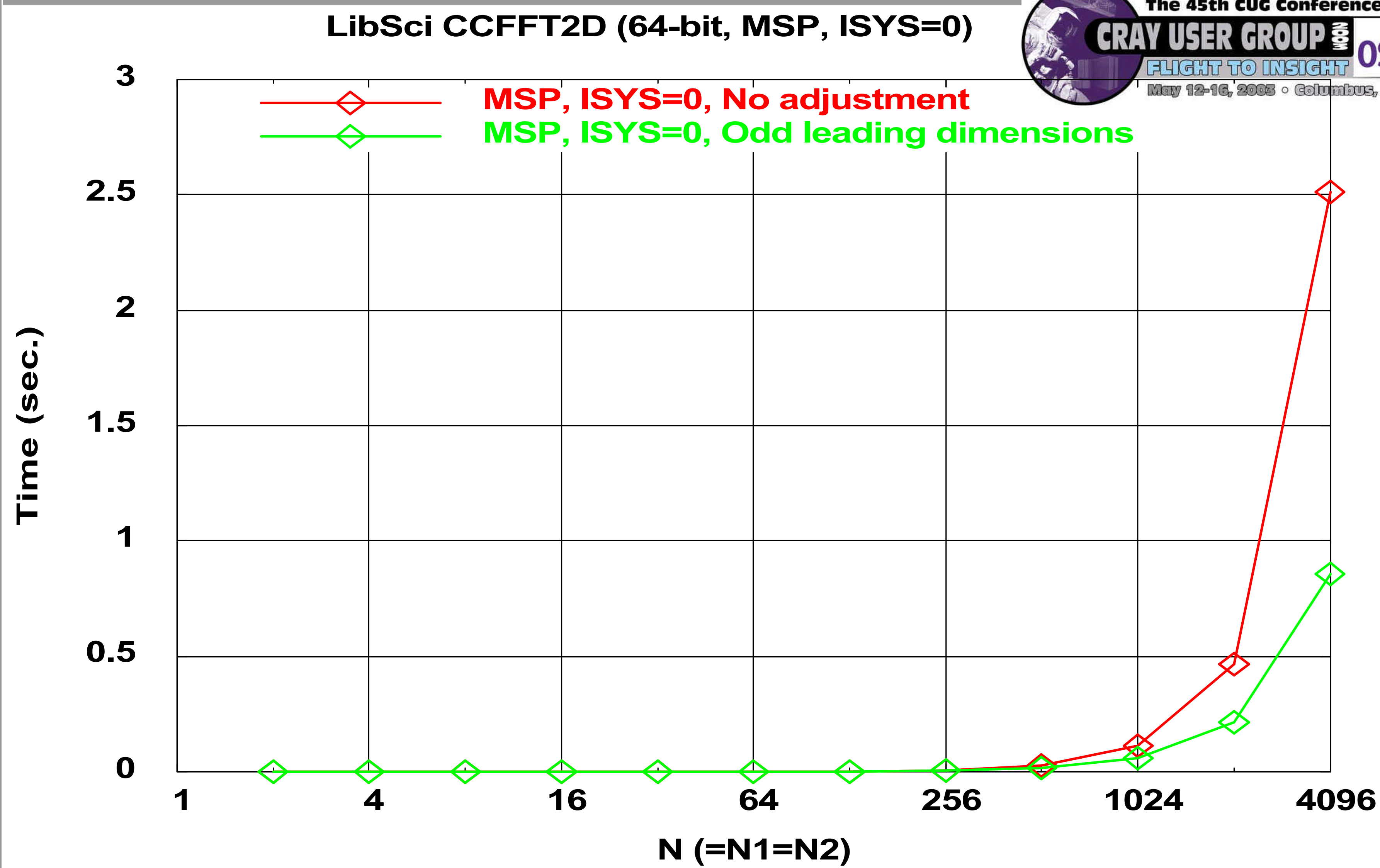
CCFFT3D (64-bit, MSP) Performance



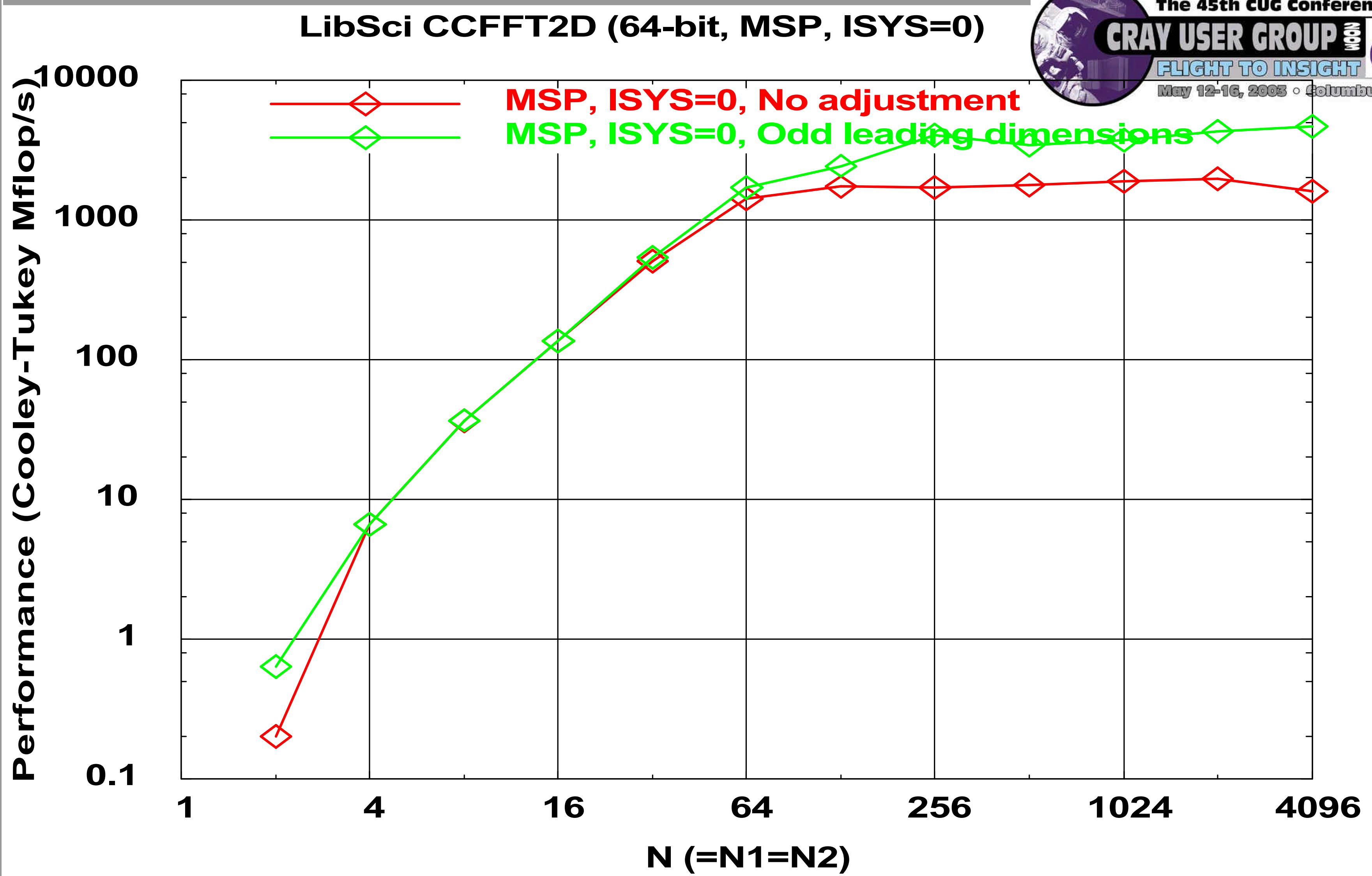
CCFFT2D (64-bit, MSP) Timings



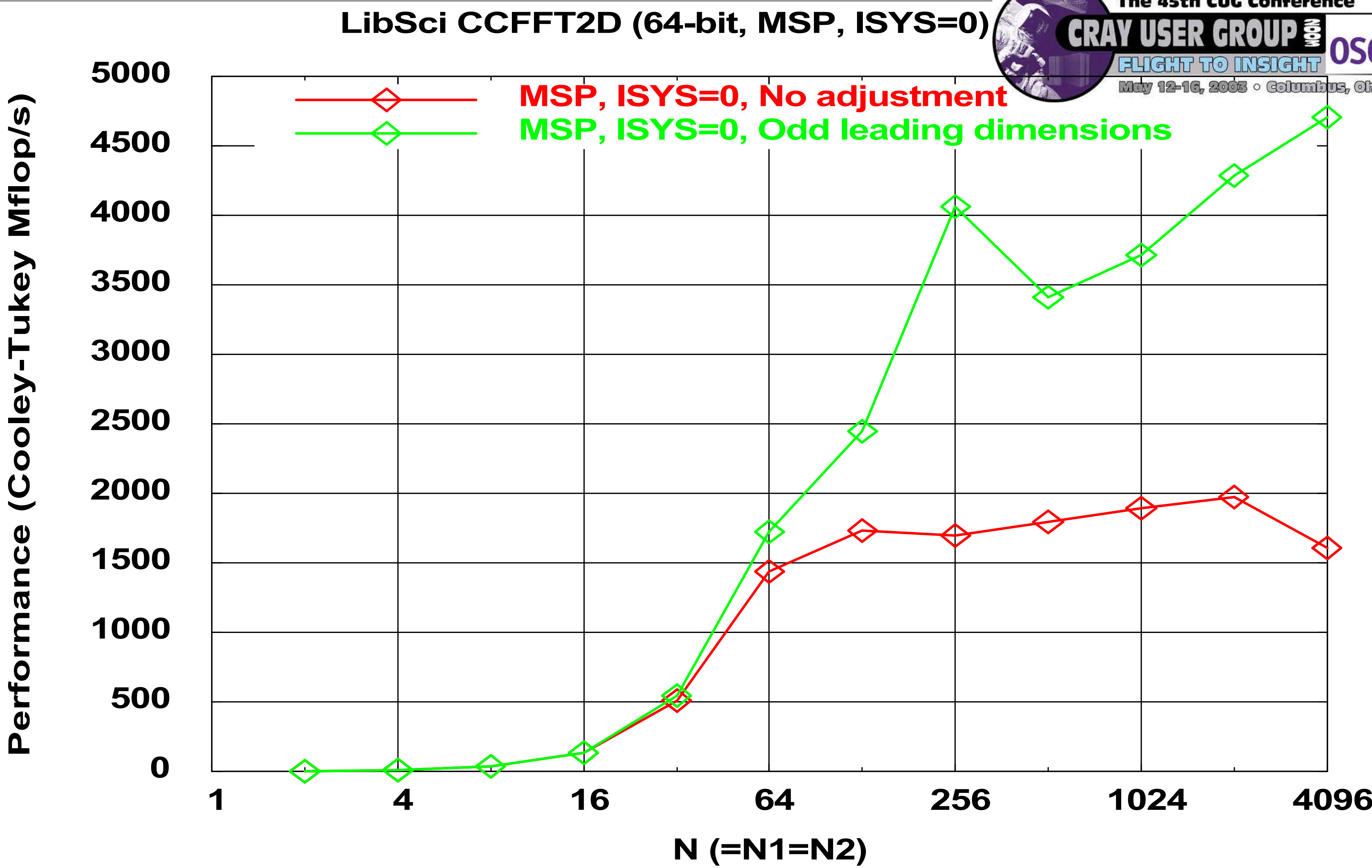
CCFFT2D (64-bit, MSP) Timings



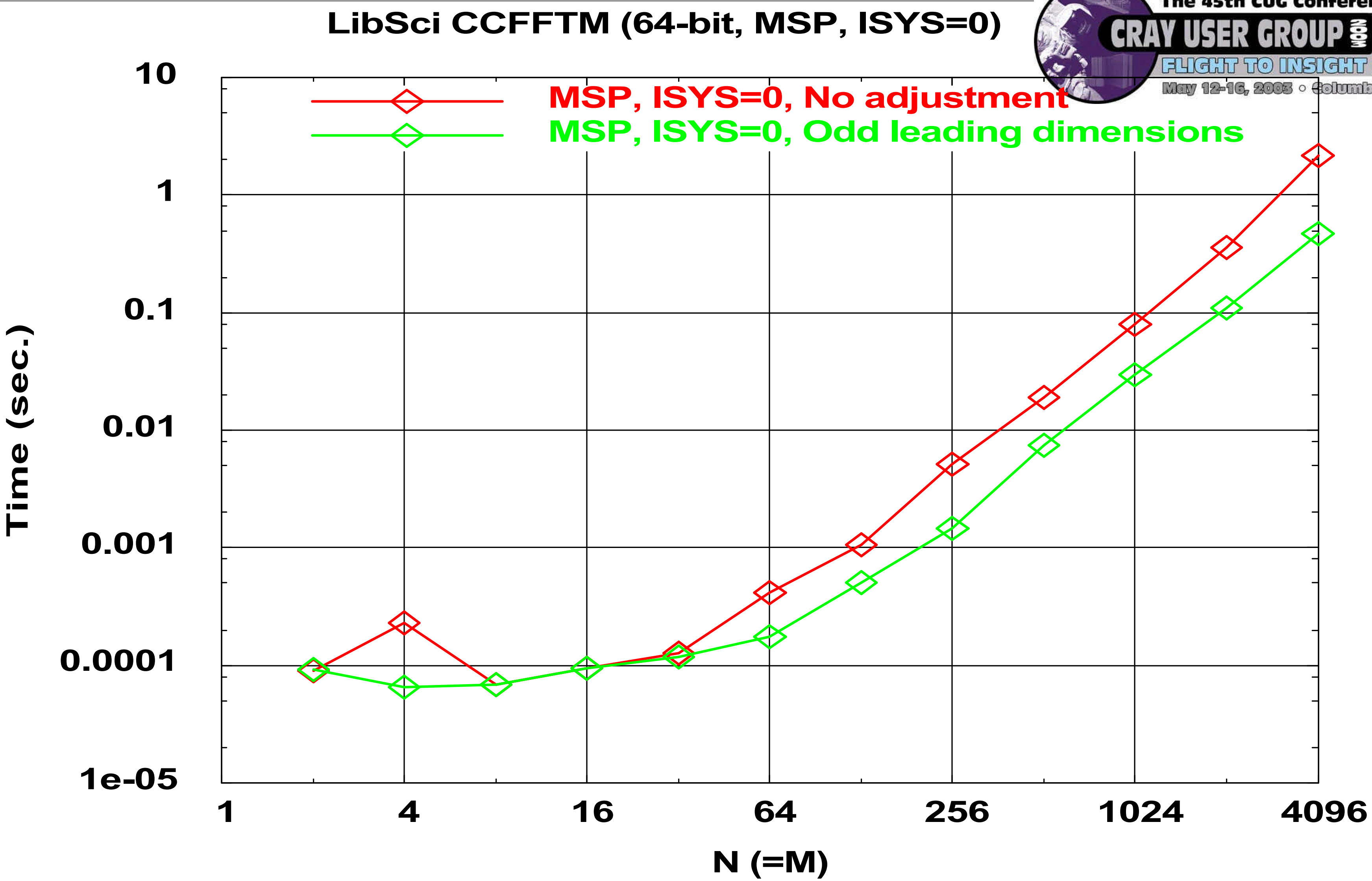
CCFFT2D (64-bit, MSP) Performance



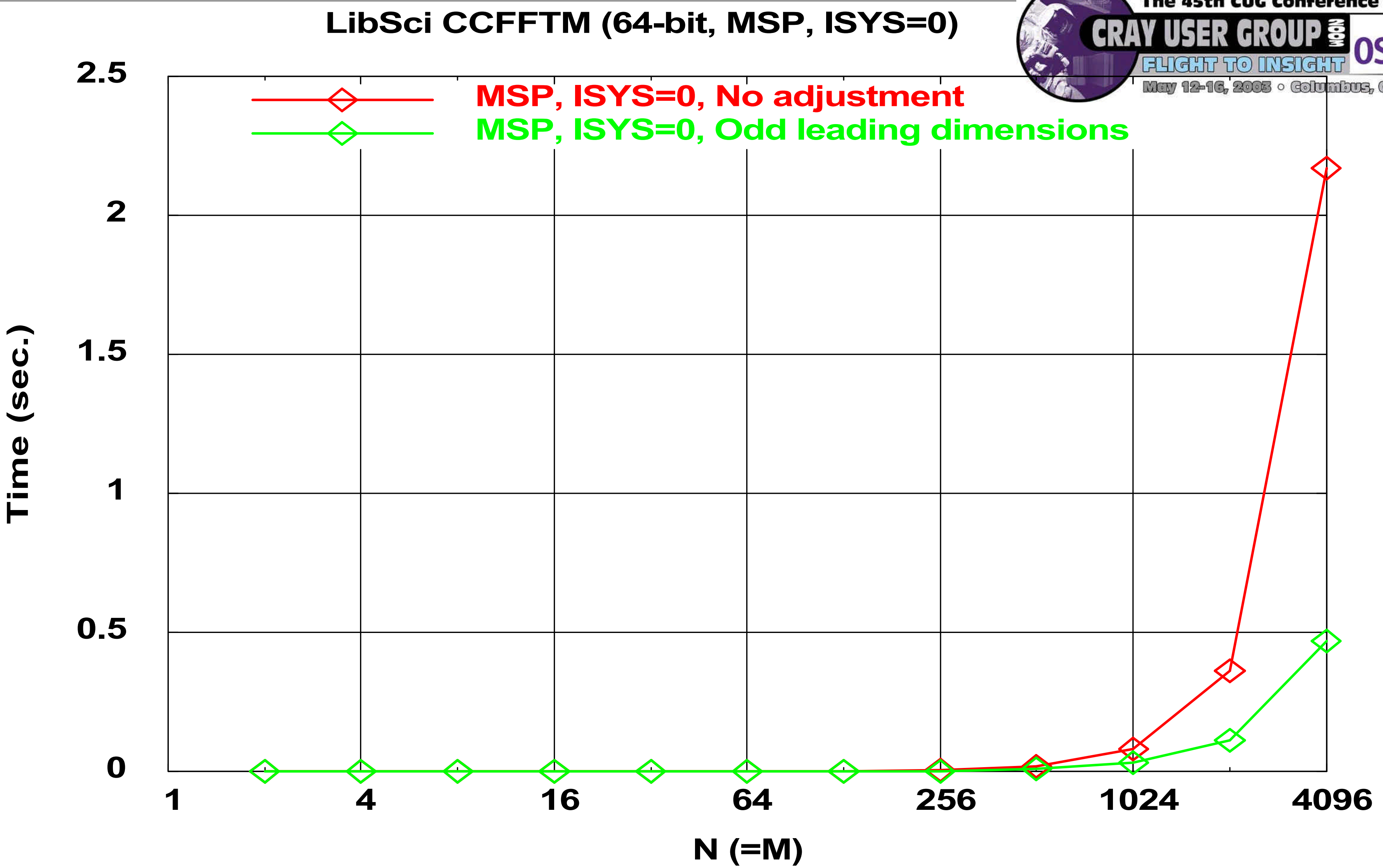
CCFFT2D (64-bit, MSP) Performance



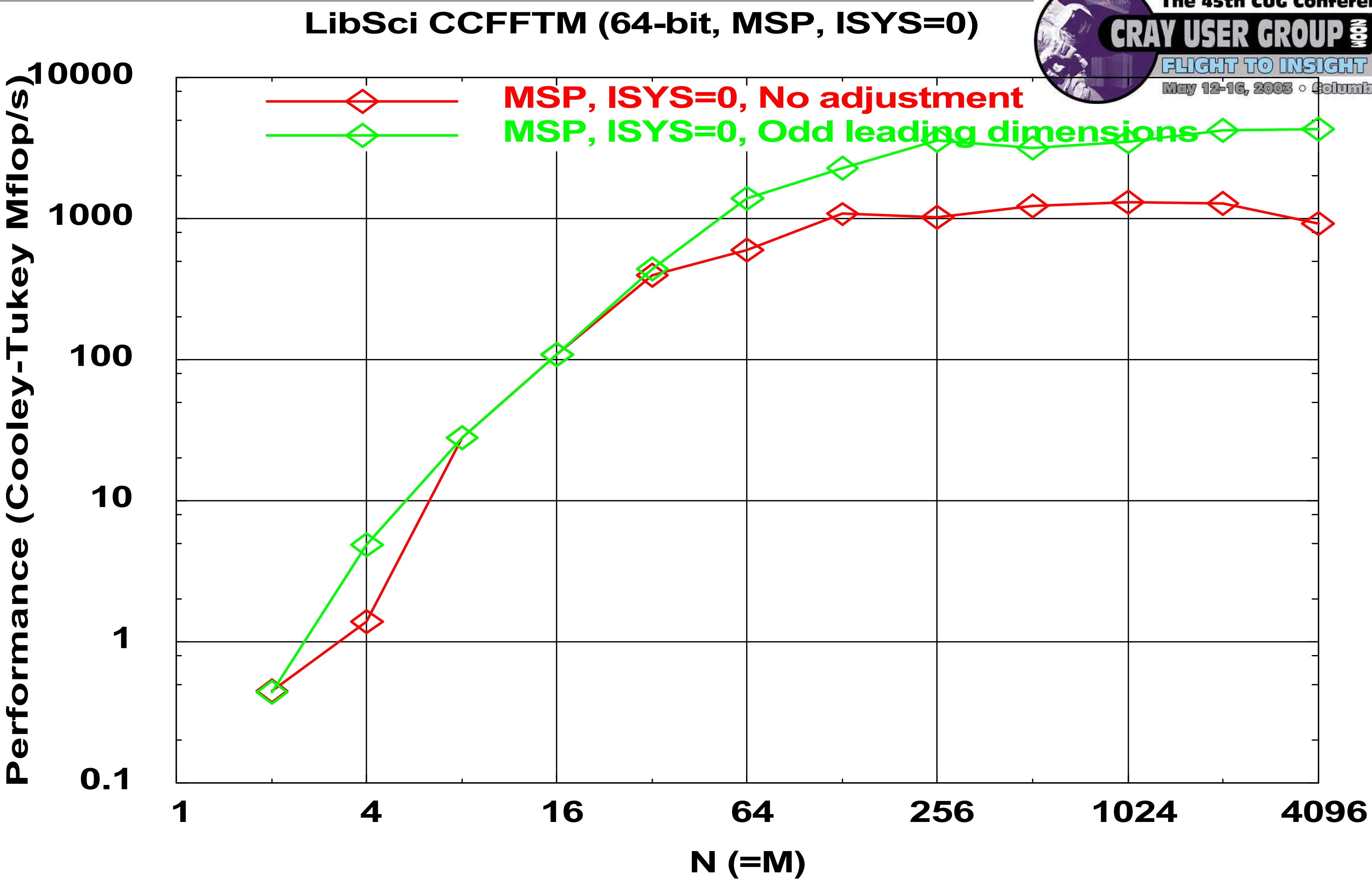
CCFFTM (64-bit, MSP) Timings



CCFFTM (64-bit, MSP) Timings

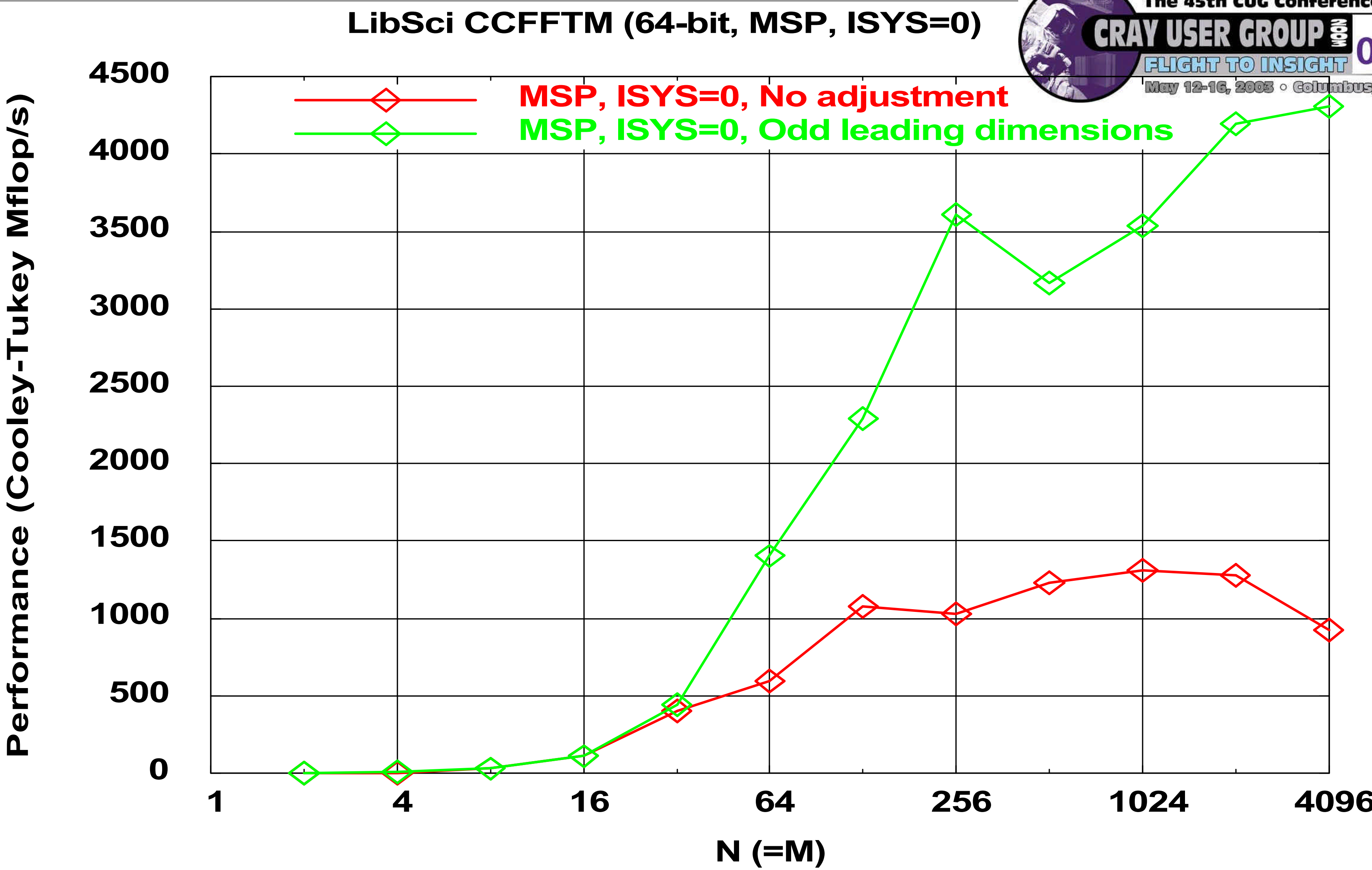


CCFFTM (64-bit, MSP) Performance



The 45th CUG Conference
CRAY USER GROUP
FLIGHT TO INSIGHT
OSC
May 12-16, 2003 ○ Columbus, Ohio

CCFFTM (64-bit, MSP) Performance



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.

Algorithm Choice for 3-D FFTs (cont.)

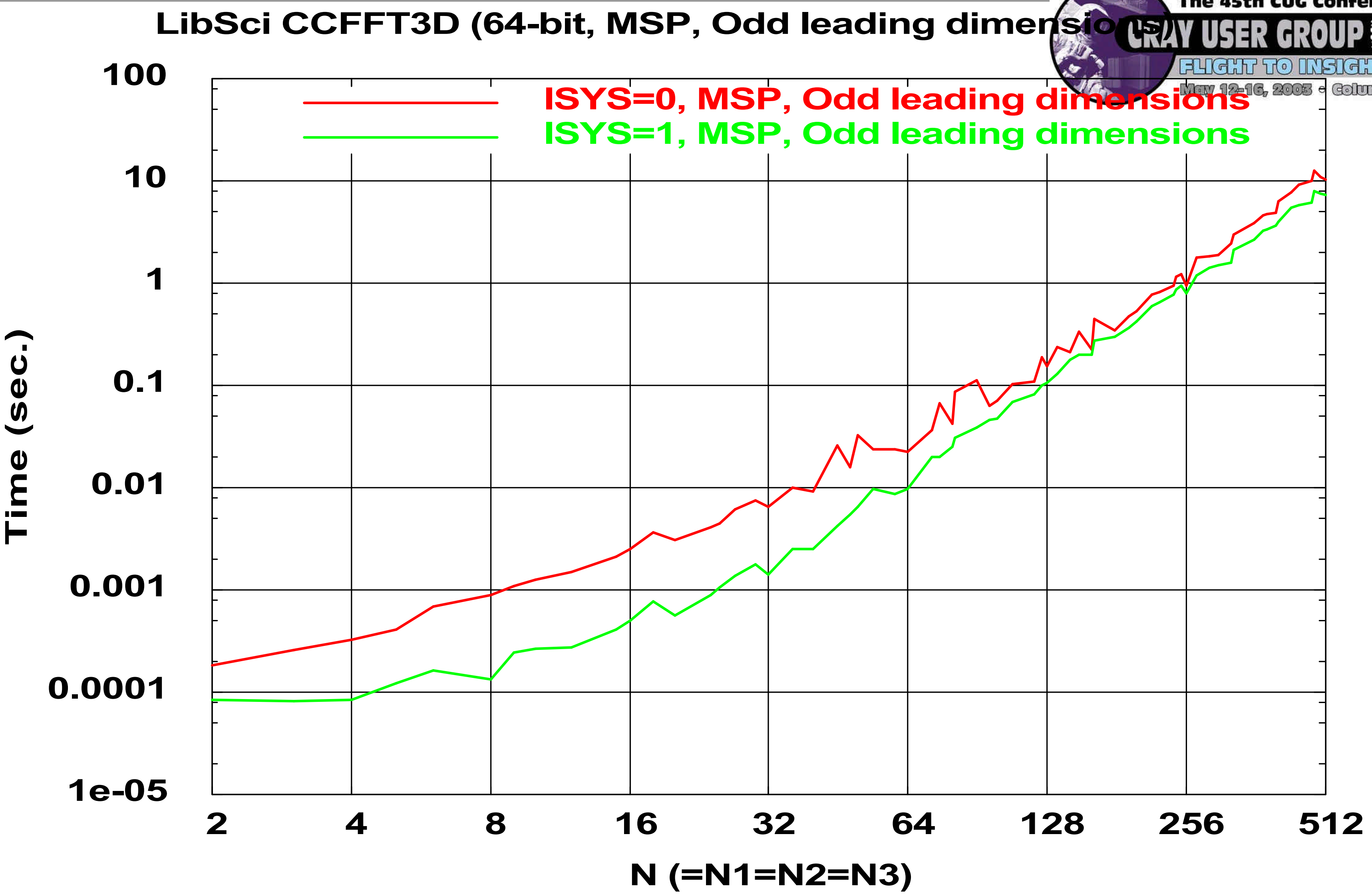
- `ISYS=1`.
  - In each of the three 1-D FFTs:
    - 1 dimension for multistreaming.
    - 1 dimension for vectorization.
    - 1 dimension for transform.
- Requires more `WORK` space than `ISYS=0`.
- Generally more performance.

Algorithm Comparisons

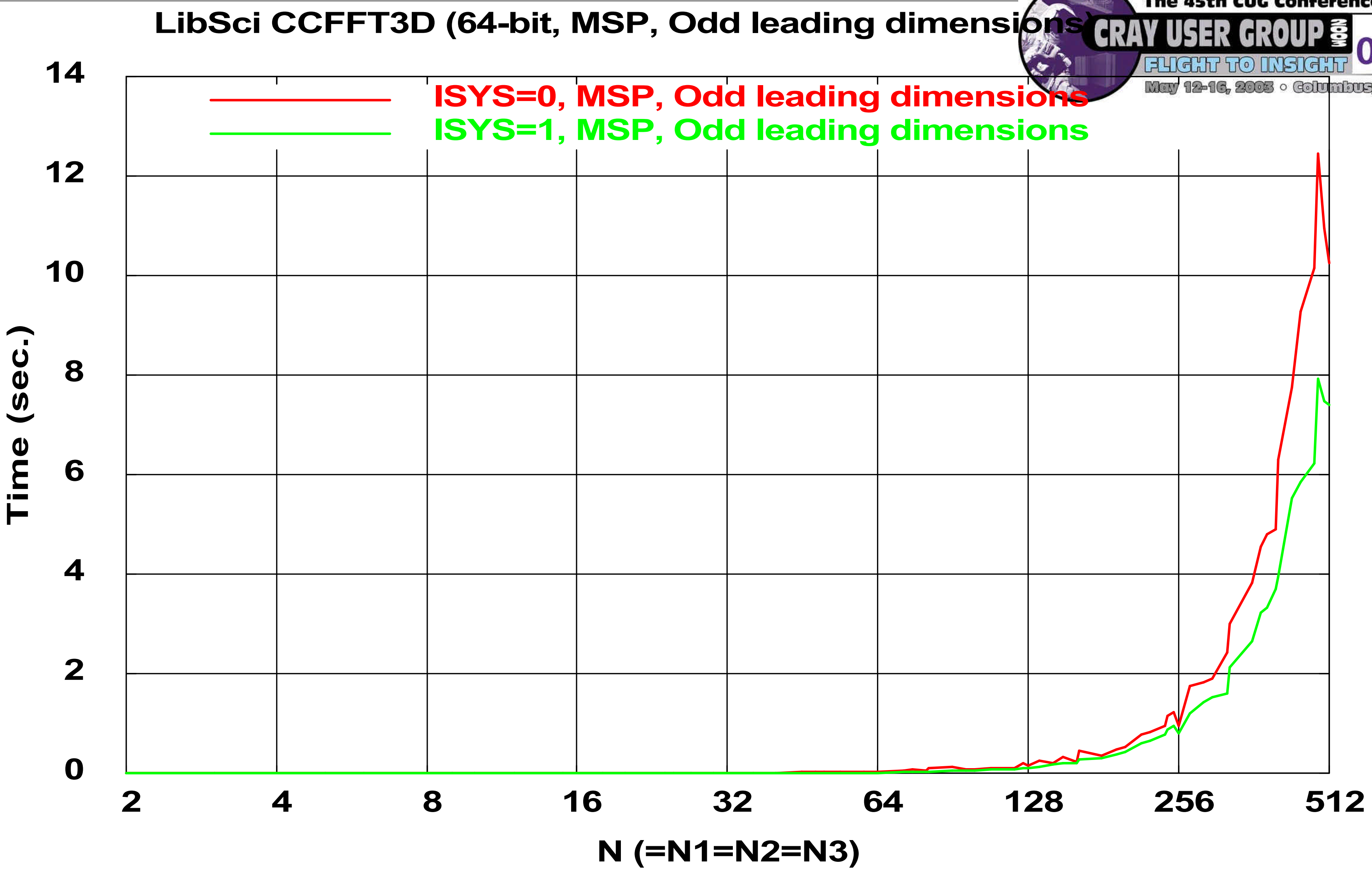
The 45th CUG Conference
CRAY USER GROUP
FLIGHT TO INSIGHT
May 12-16, 2003 ◦ Columbus, Ohio

- 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.

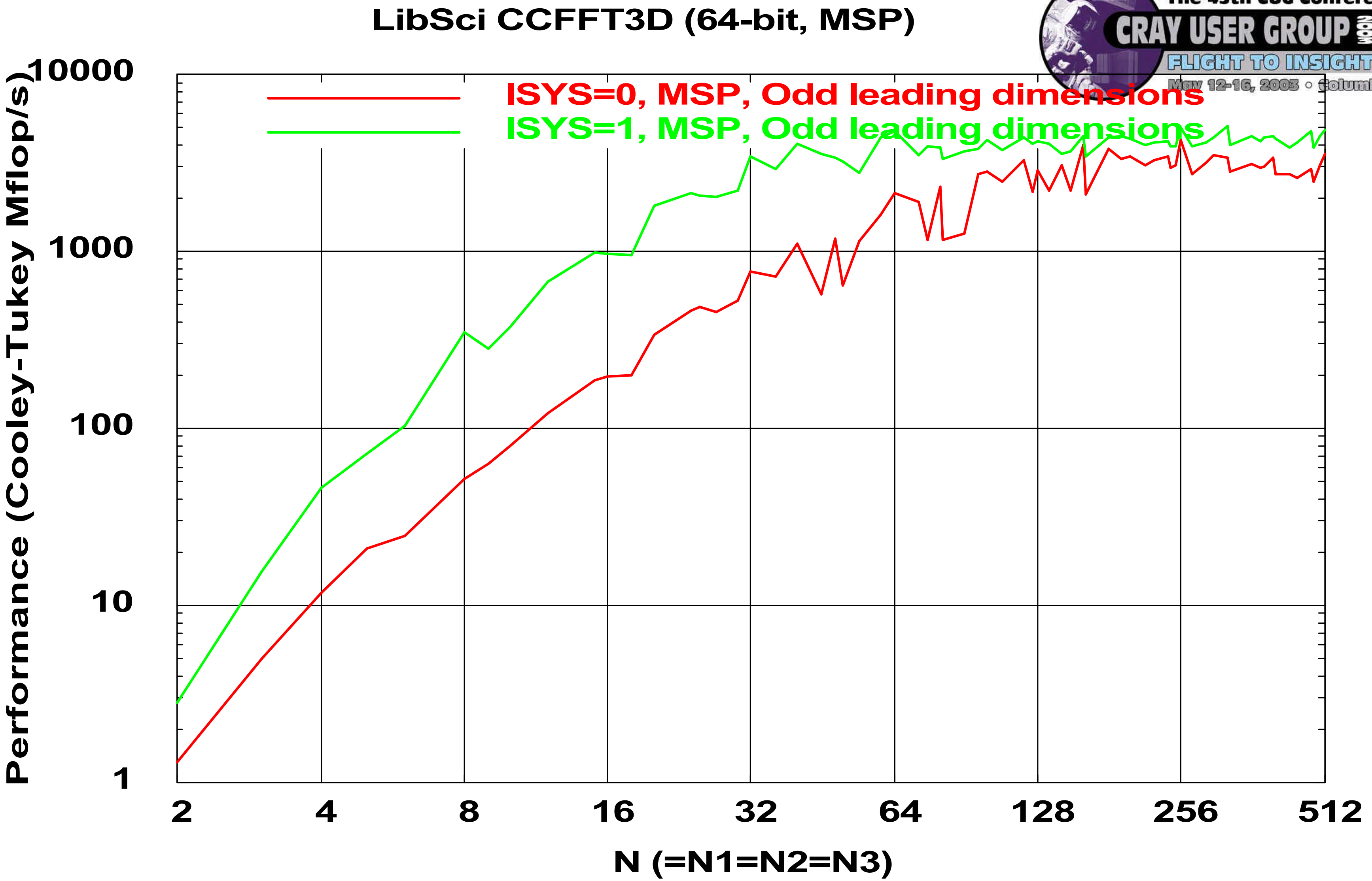
CCFFT3D (64-bit, MSP) Timings



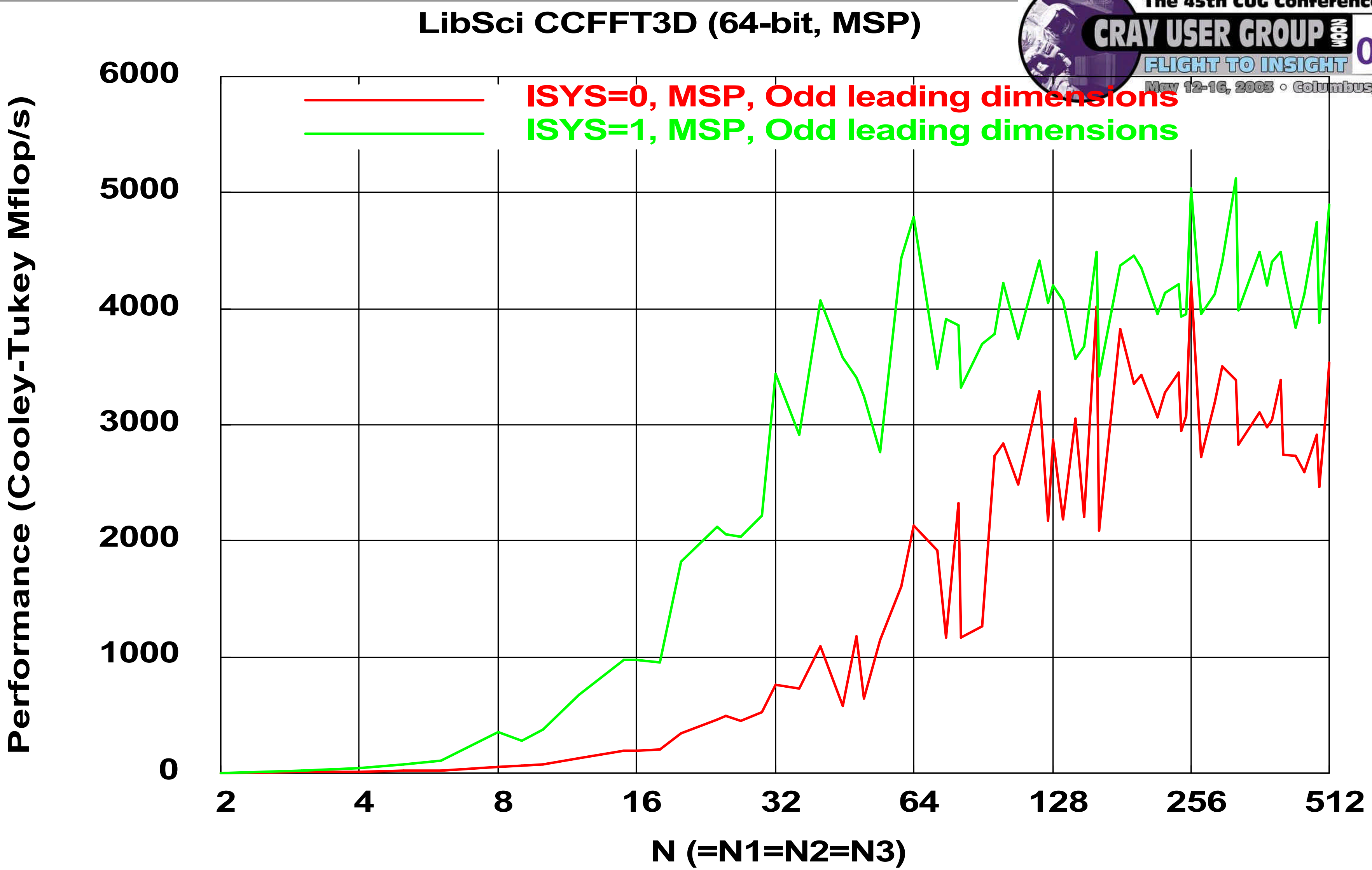
CCFFT3D (64-bit, MSP) Timings



CCFFT3D (64-bit, MSP) Performance



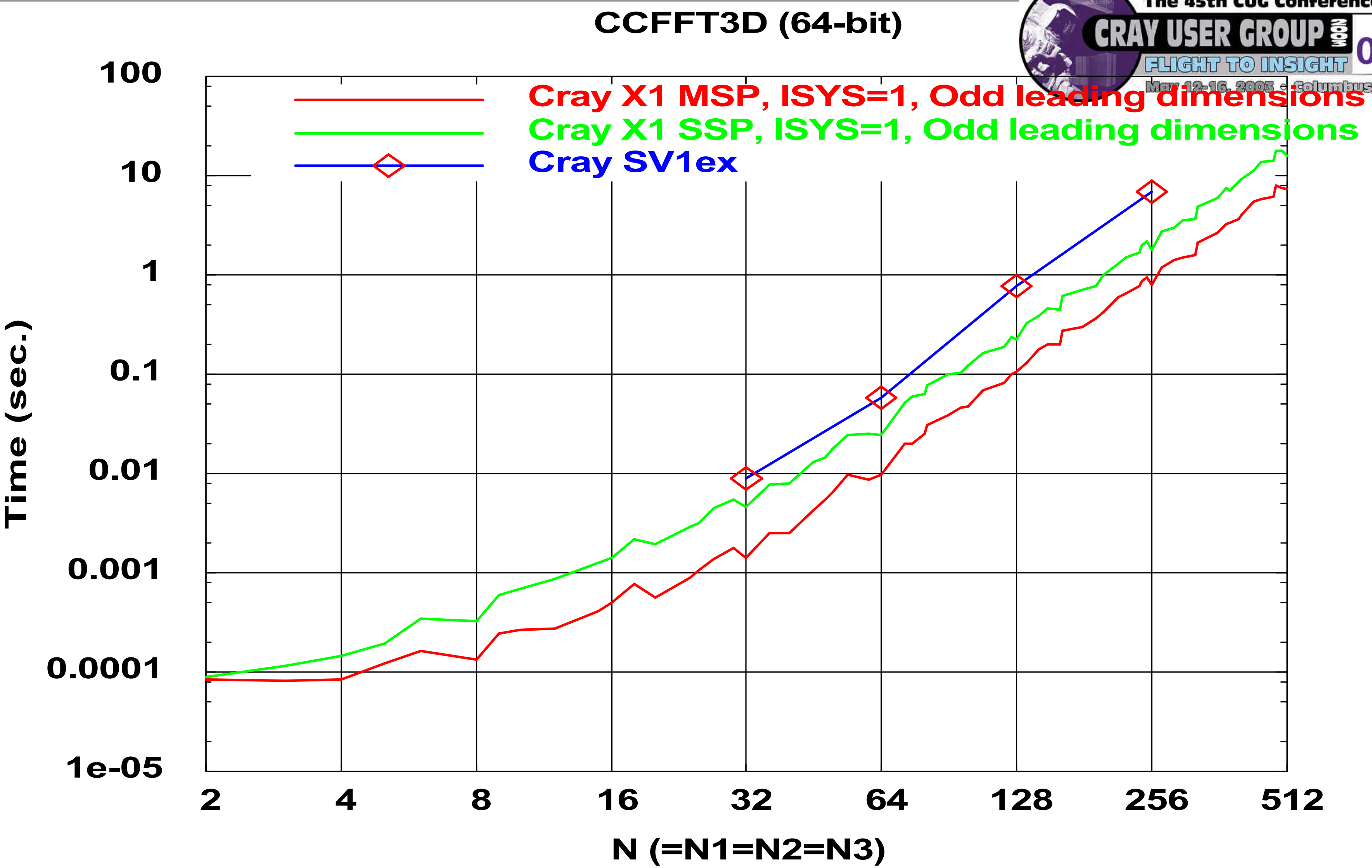
CCFFT3D (64-bit, MSP) Performance



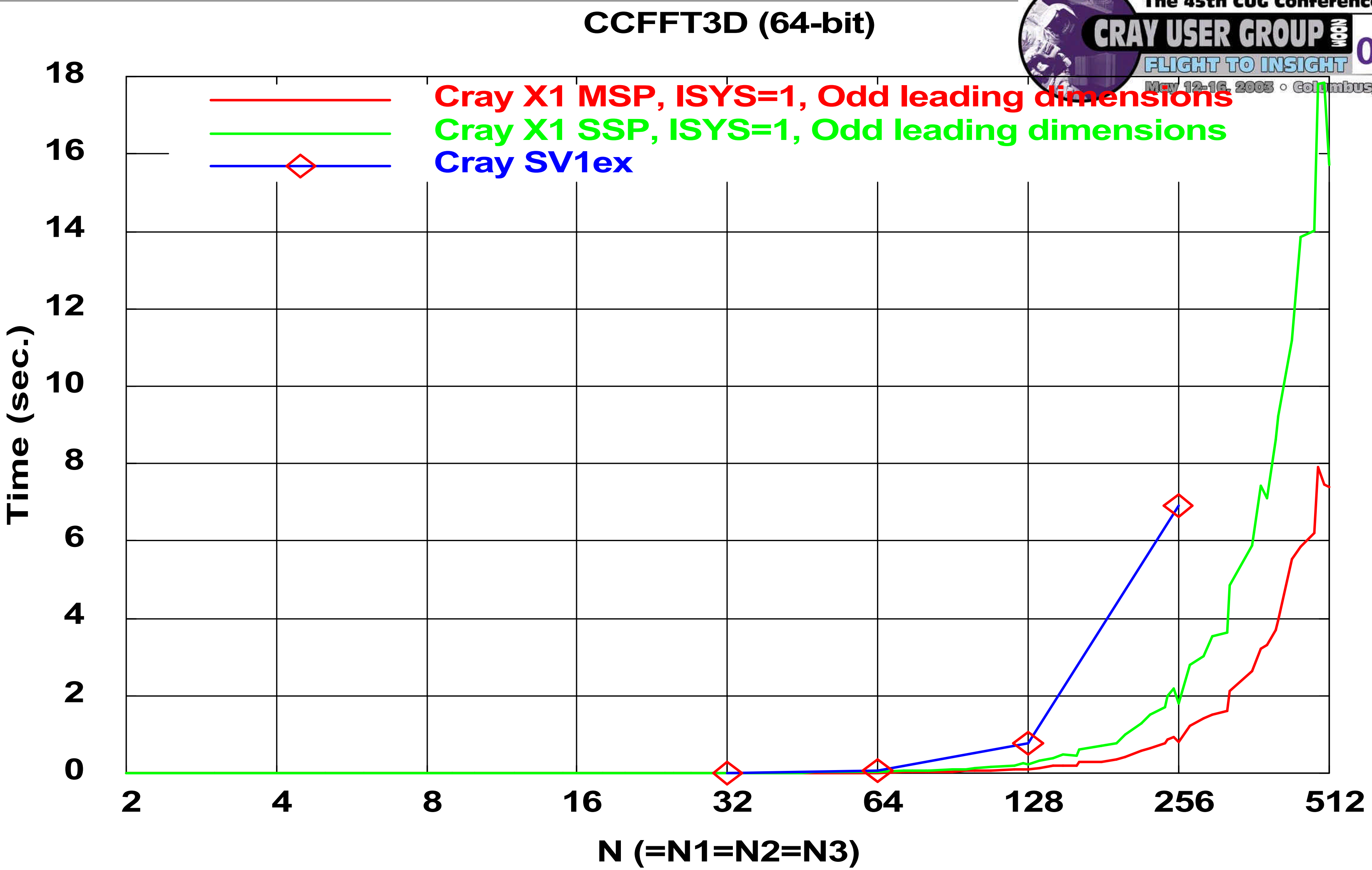
(IV) Performance & Timings

- 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.
- ISYS = 0 vs. ISYS = 1.
- No lengths containing factors not powers of 2, 3, and 5.

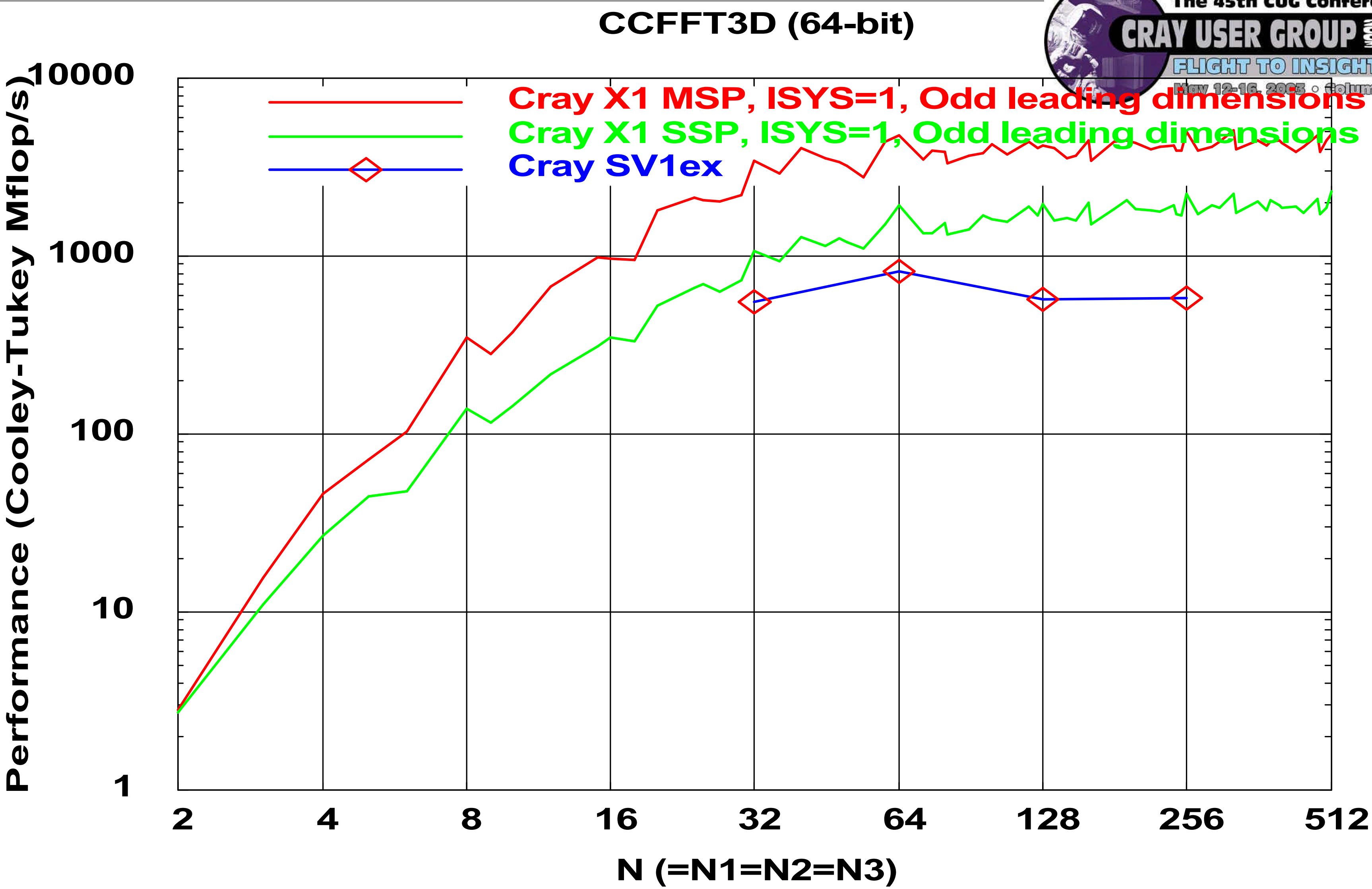
CCFFT3D (64-bit) Timings



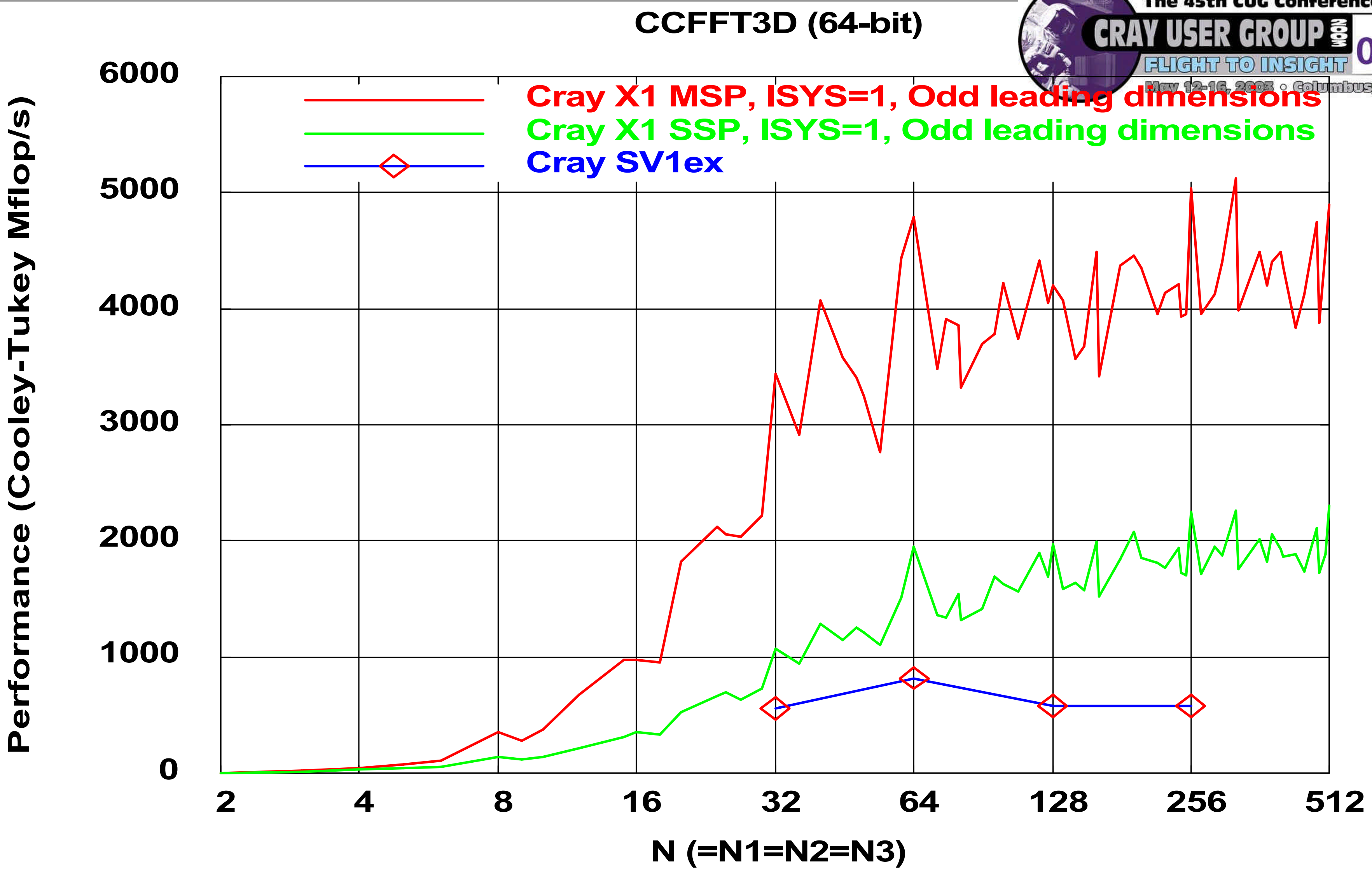
CCFFT3D (64-bit) Timings



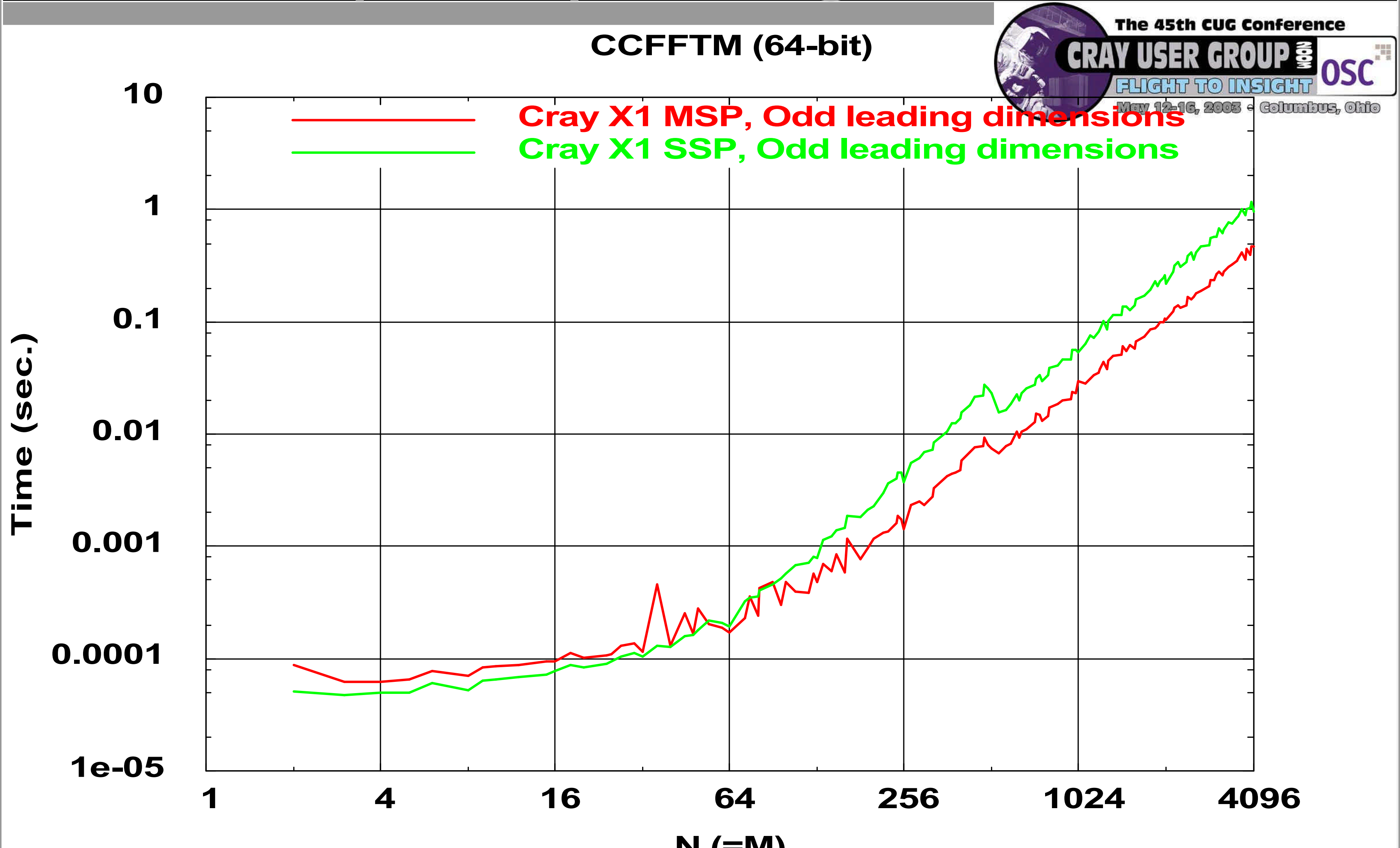
CCFFT3D (64-bit) Performance



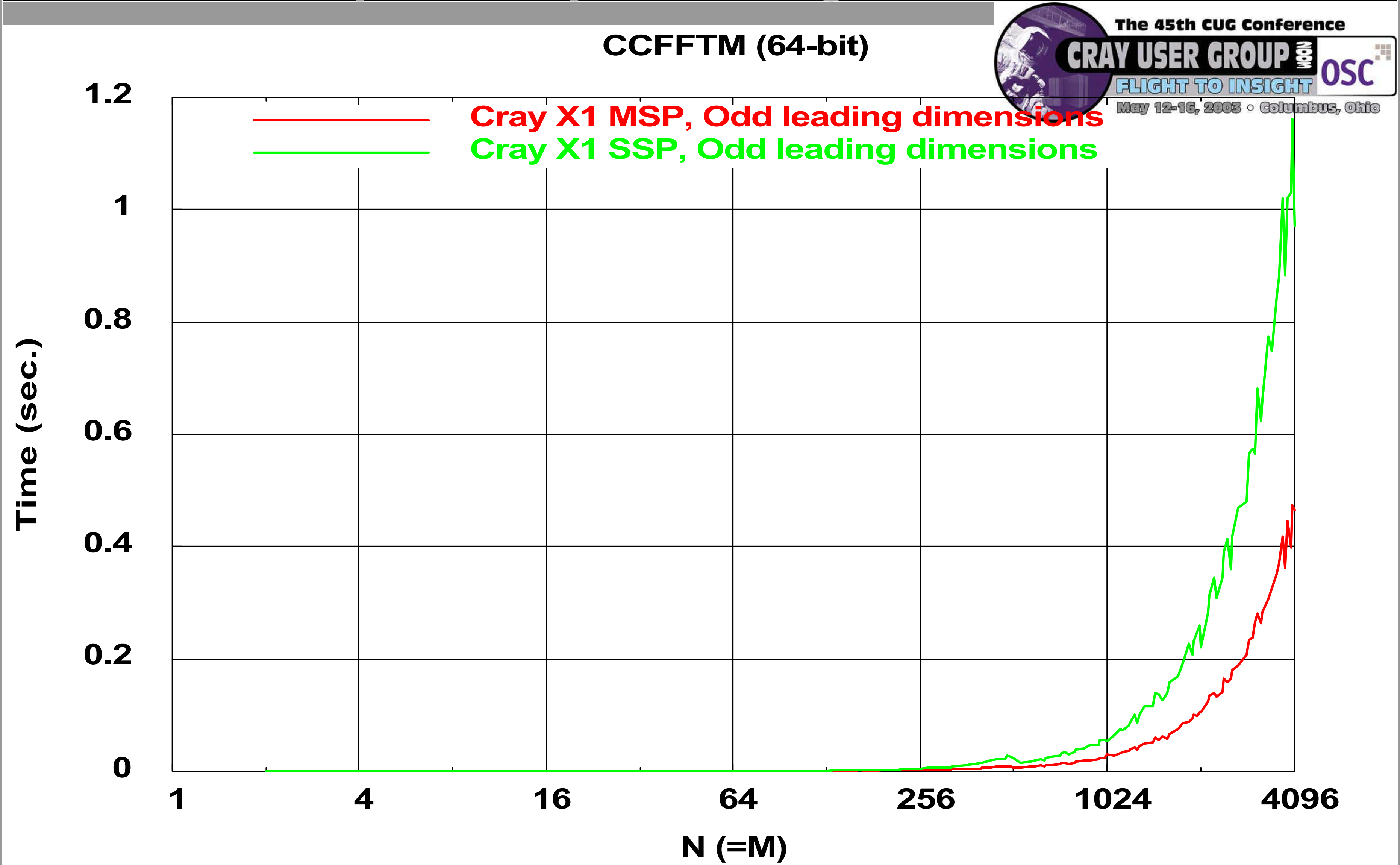
CCFFT3D (64-bit) Performance



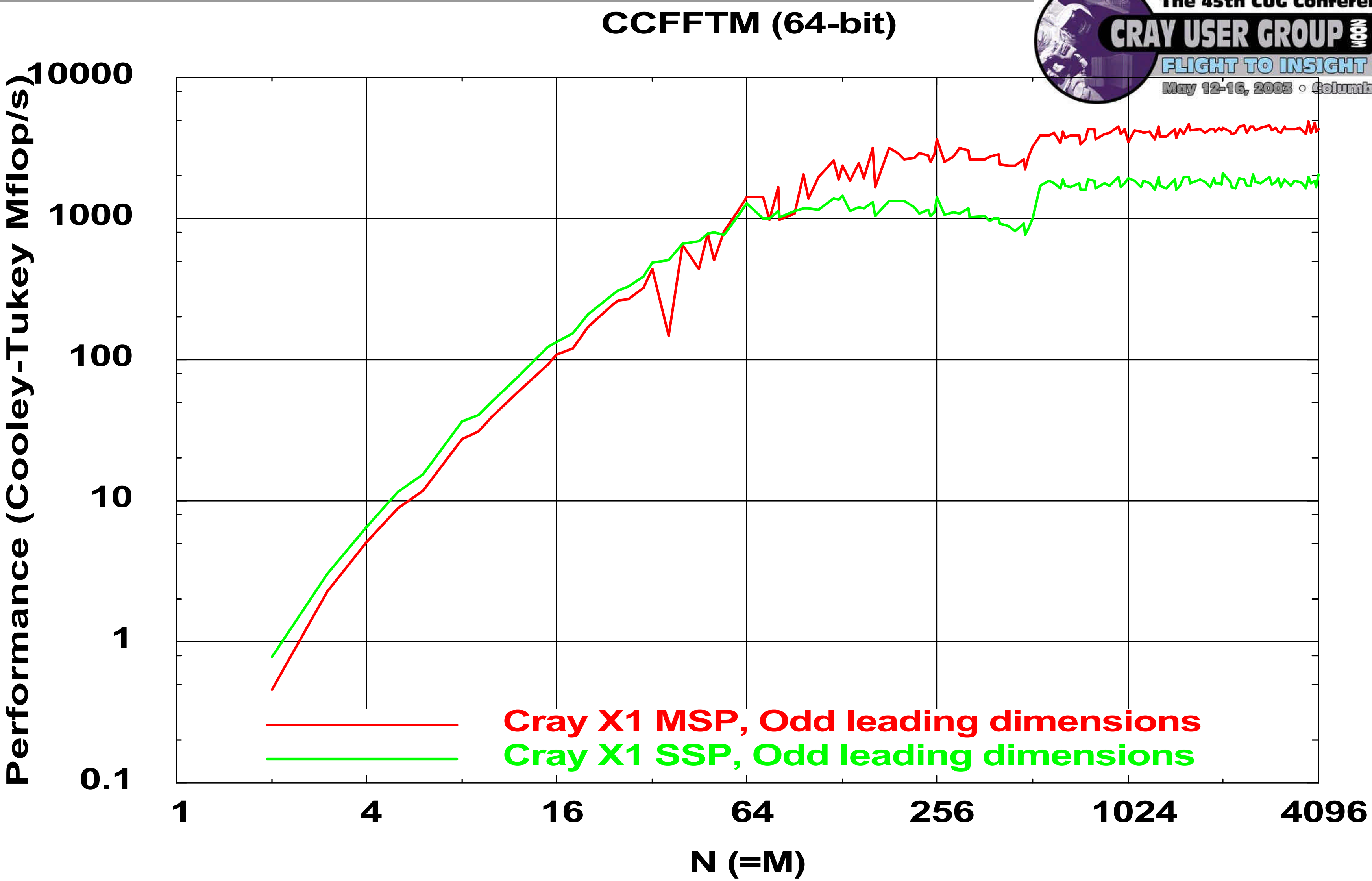
CCFFTM (64-bit) Timings



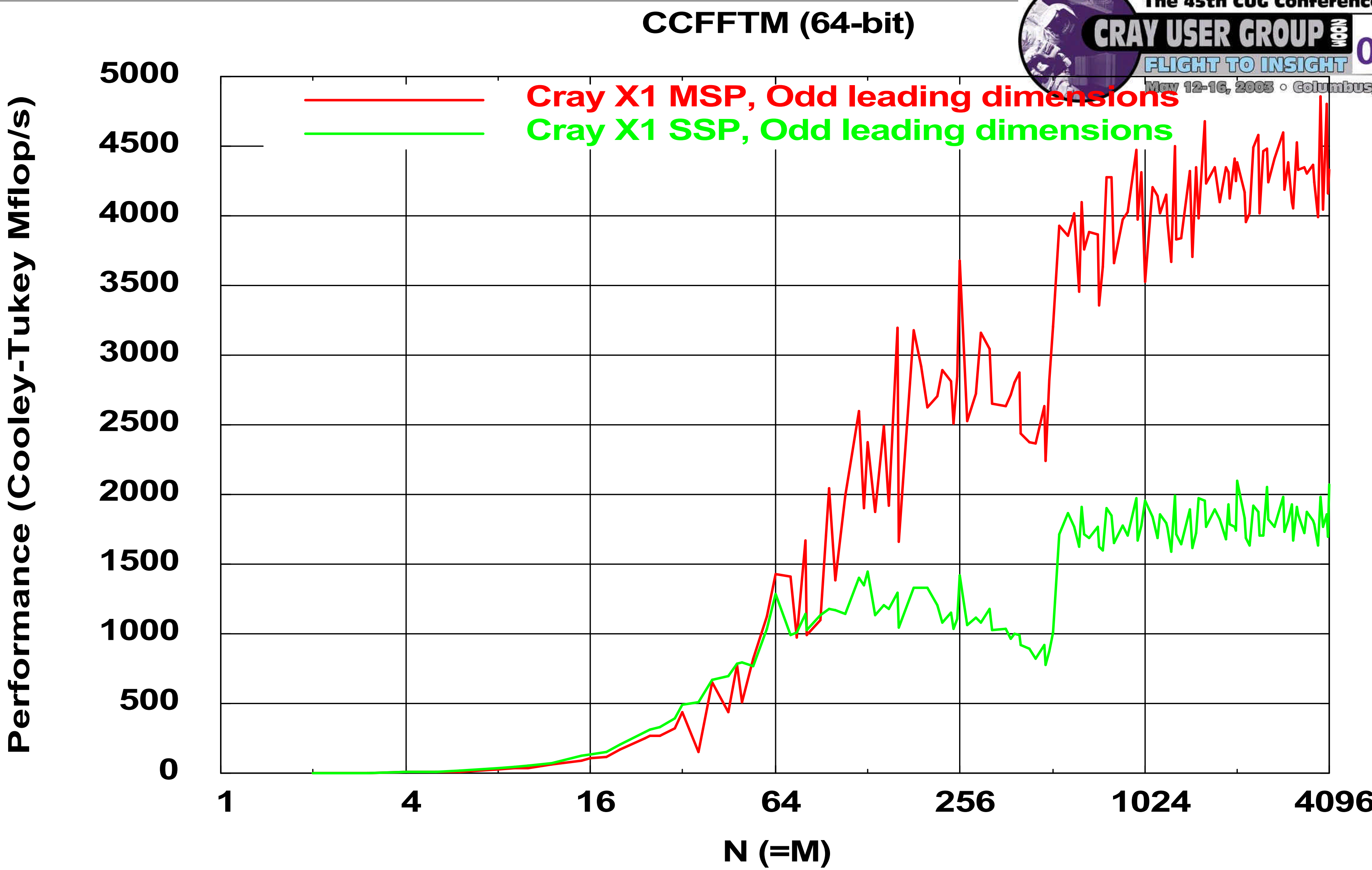
CCFFTM (64-bit) Timings



CCFFTM (64-bit) Performance

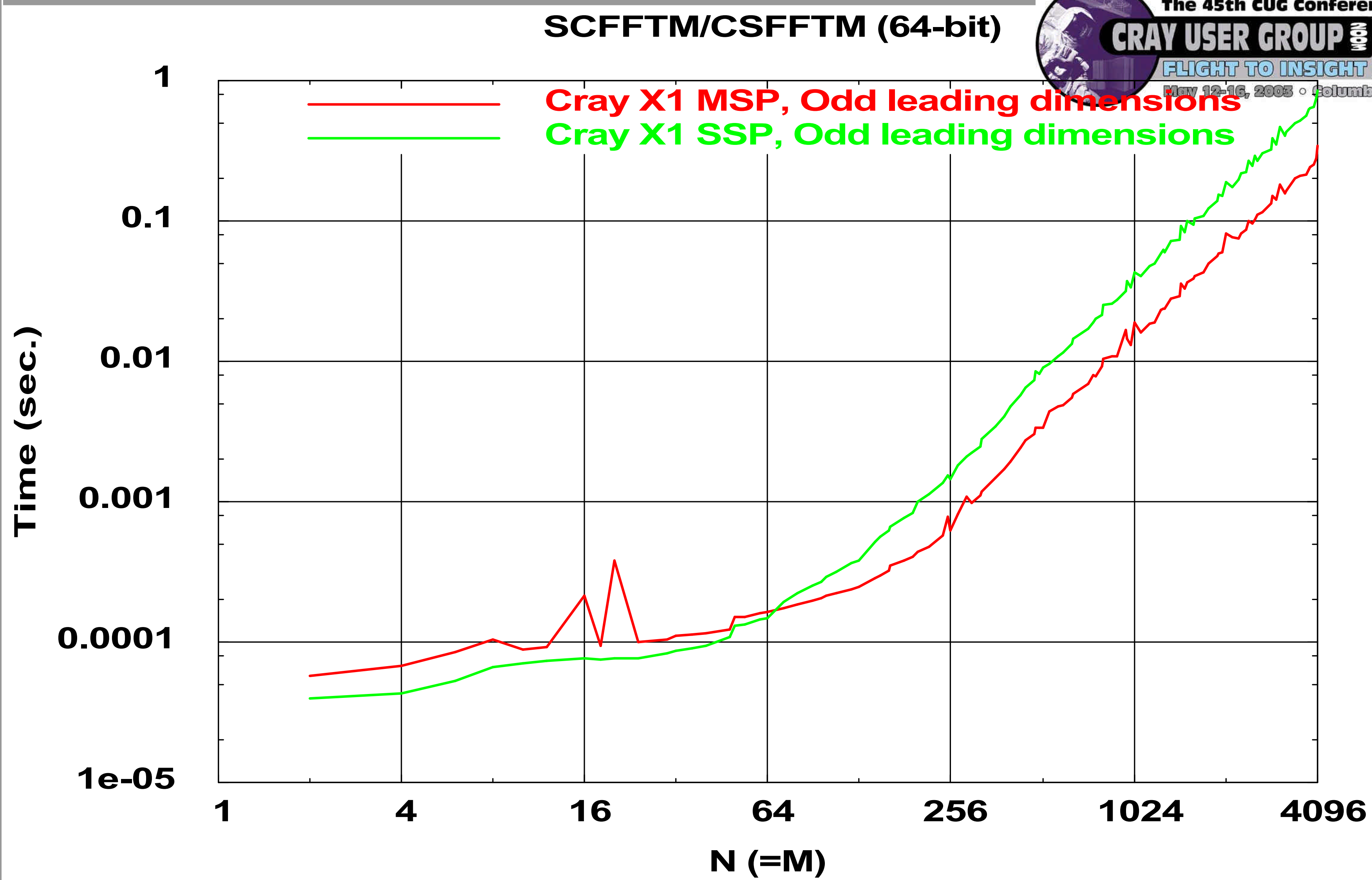


CCFFTM (64-bit) Performance

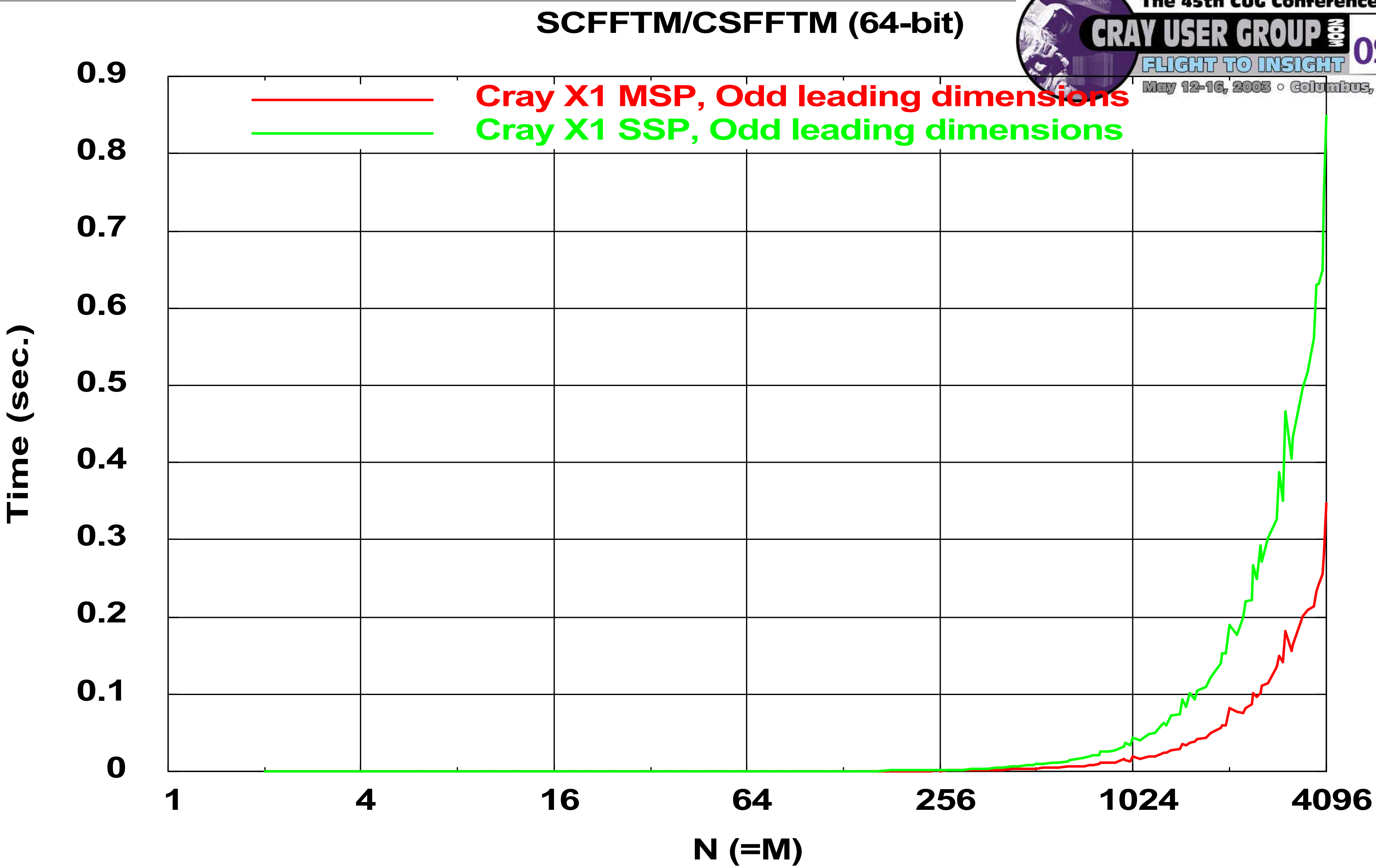


The 45th CUG Conference
CRAY USER GROUP
FLIGHT TO INSIGHT
OSC
May 12-16, 2003 ○ Columbus, Ohio

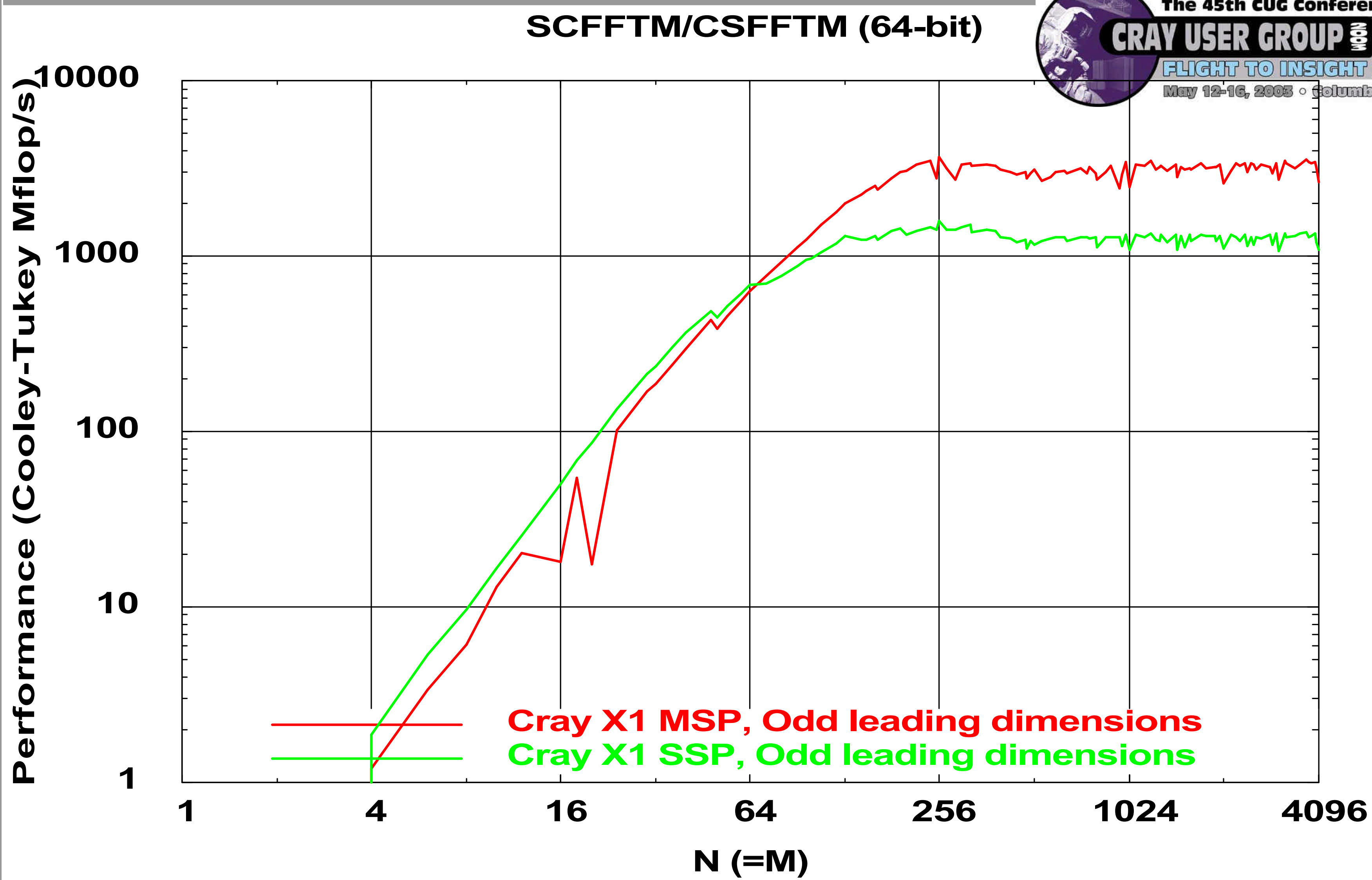
SC/CSFFTM (64-bit) Timings



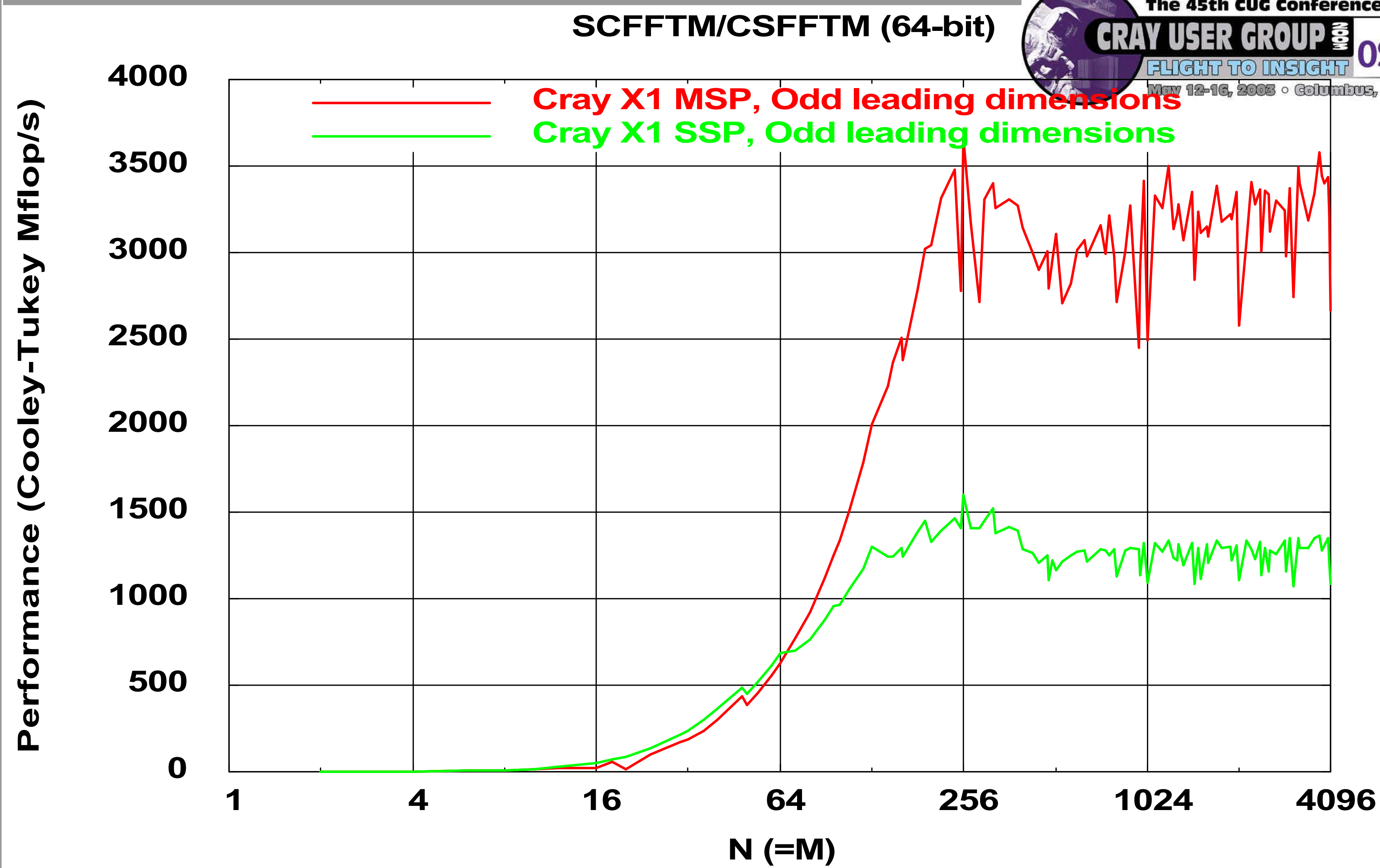
SC/CSFFTM (64-bit) Timings



SC/CSFFTM (64-bit) Performance

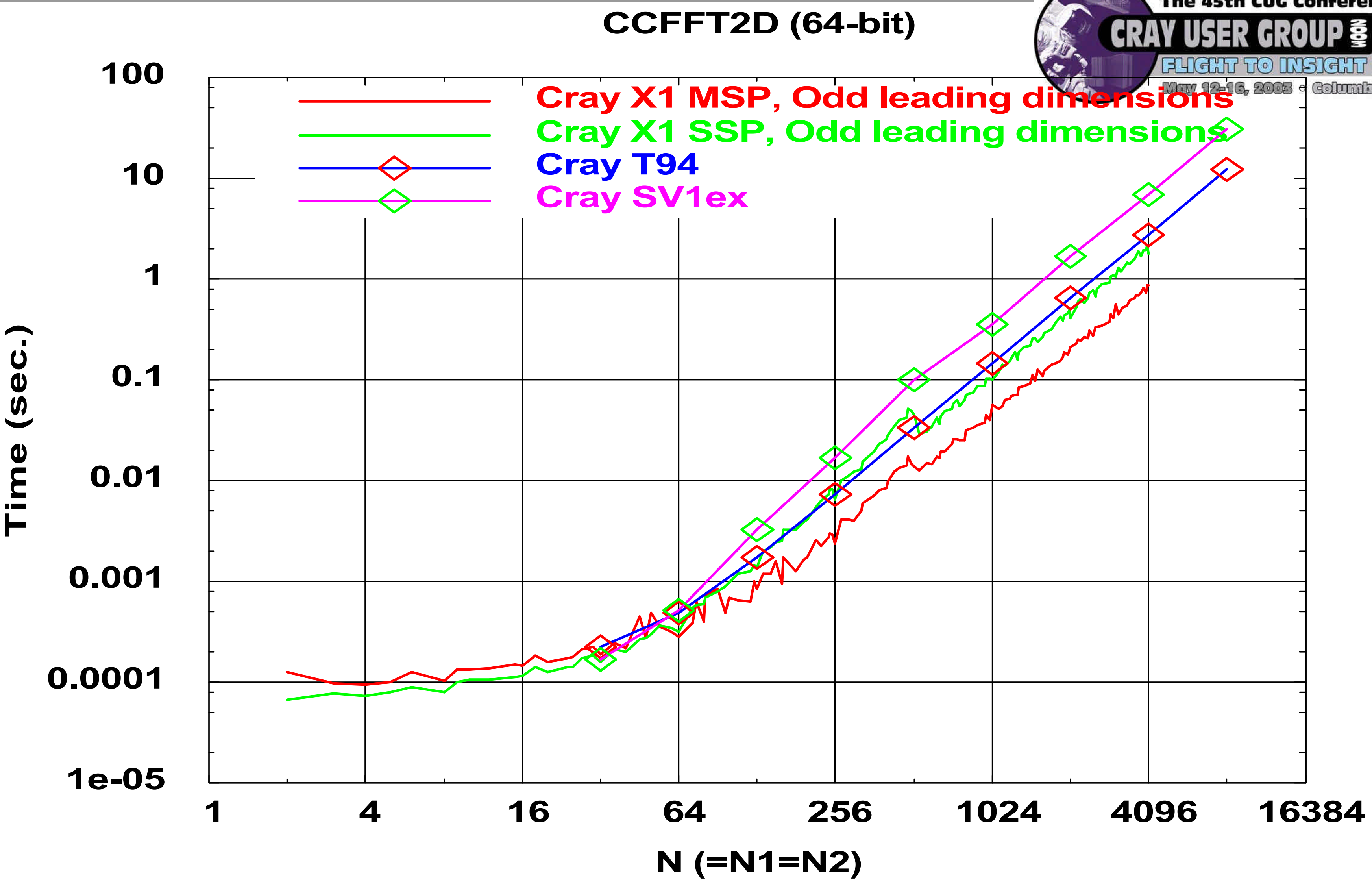


SC/CSFFTM (64-bit) Performance

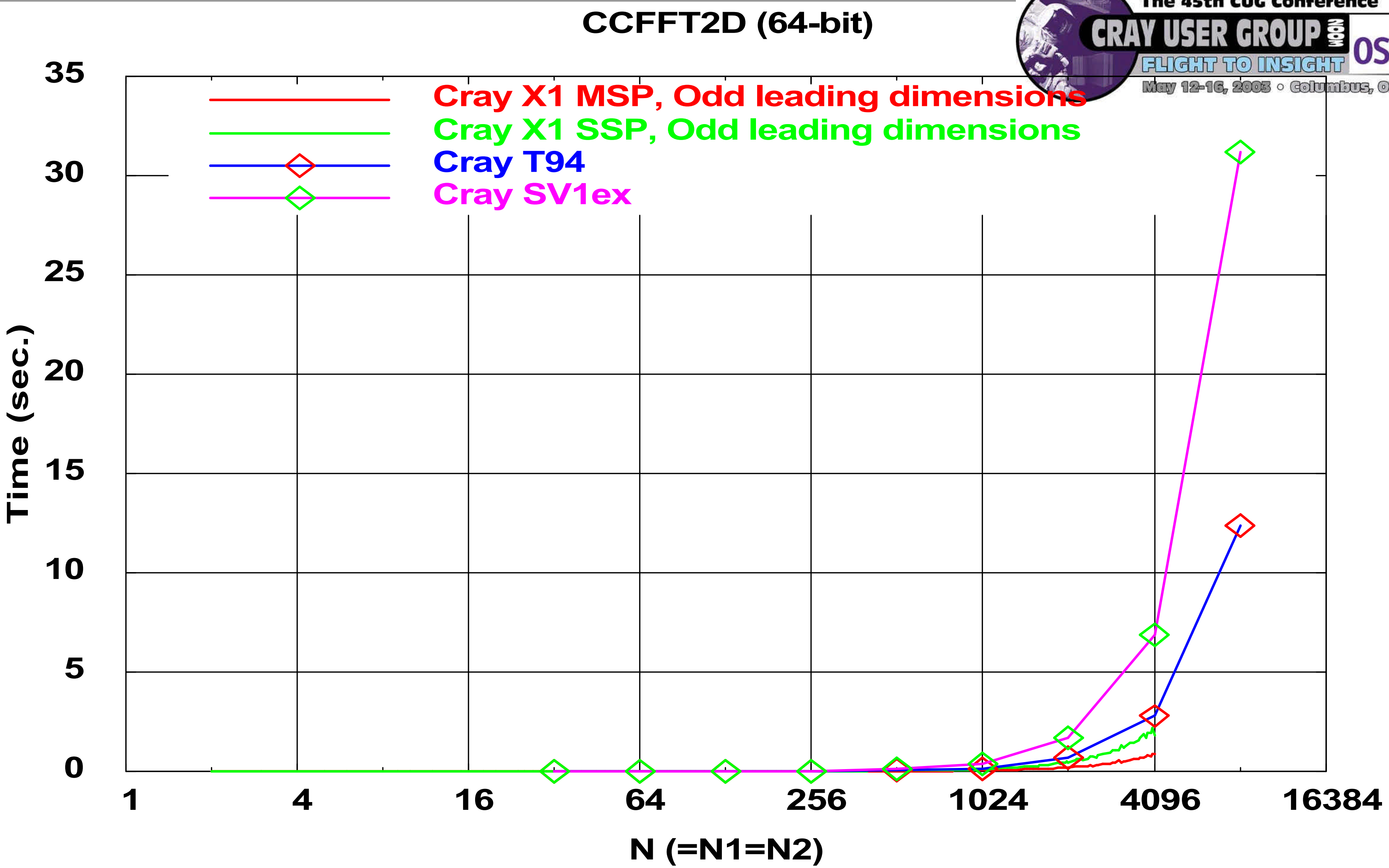


The 45th CUG Conference
CRAY USER GROUP
FLIGHT TO INSIGHT
May 12-16, 2003 ◦ Columbus, Ohio
OSC

CCFFT2D (64-bit) Timings

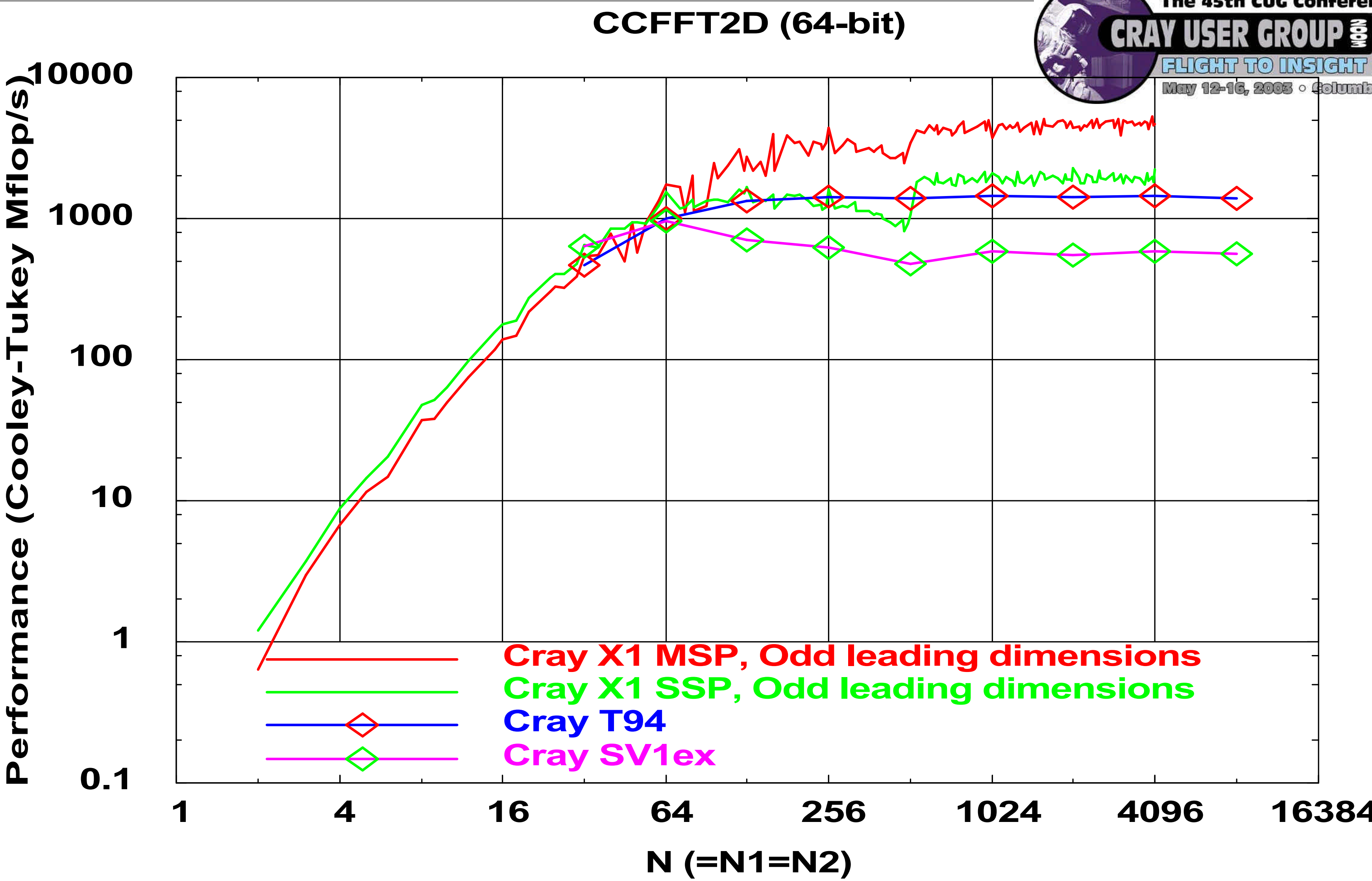


CCFFT2D (64-bit) Timings

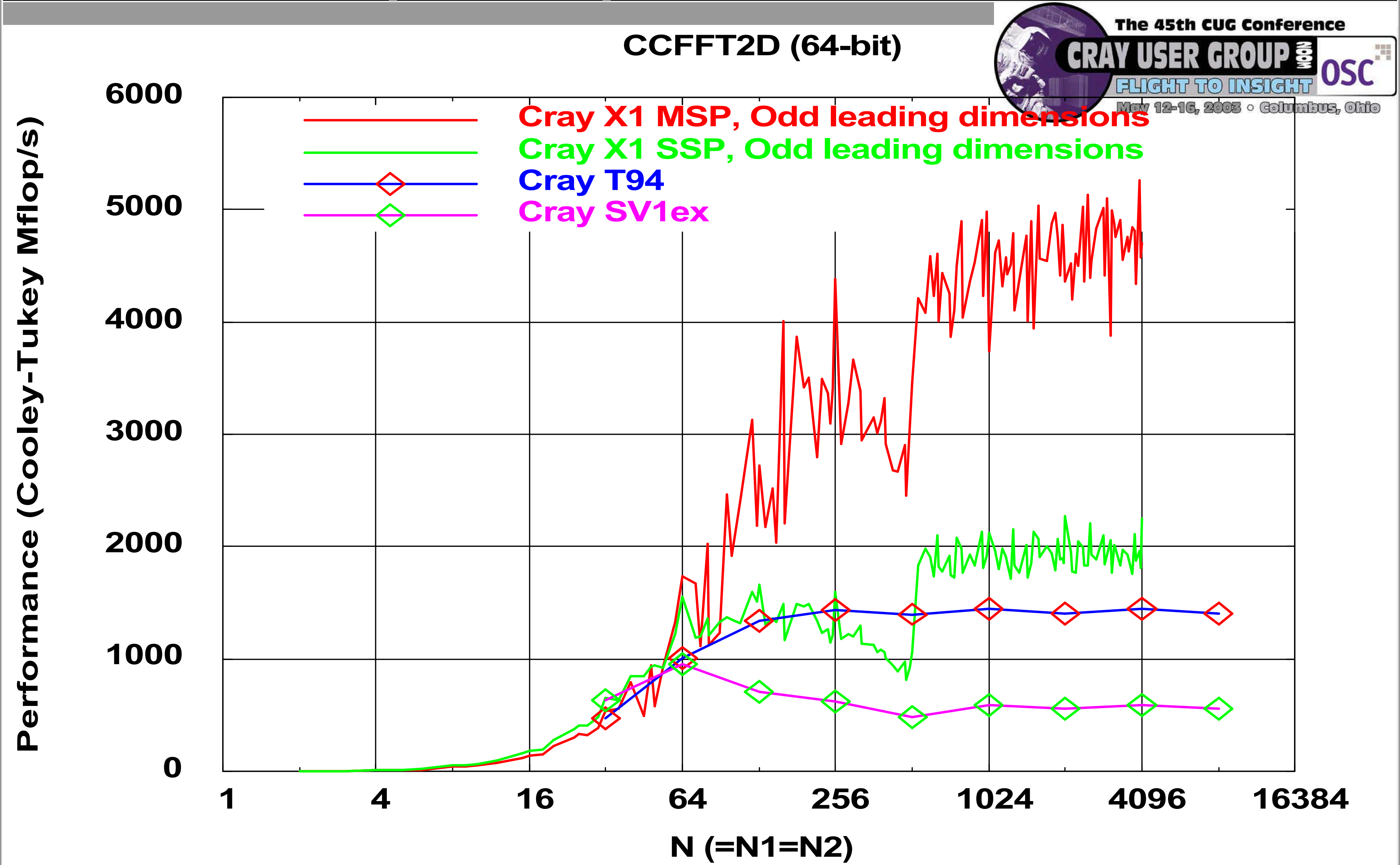


CCFFT2D (64-bit) Performance

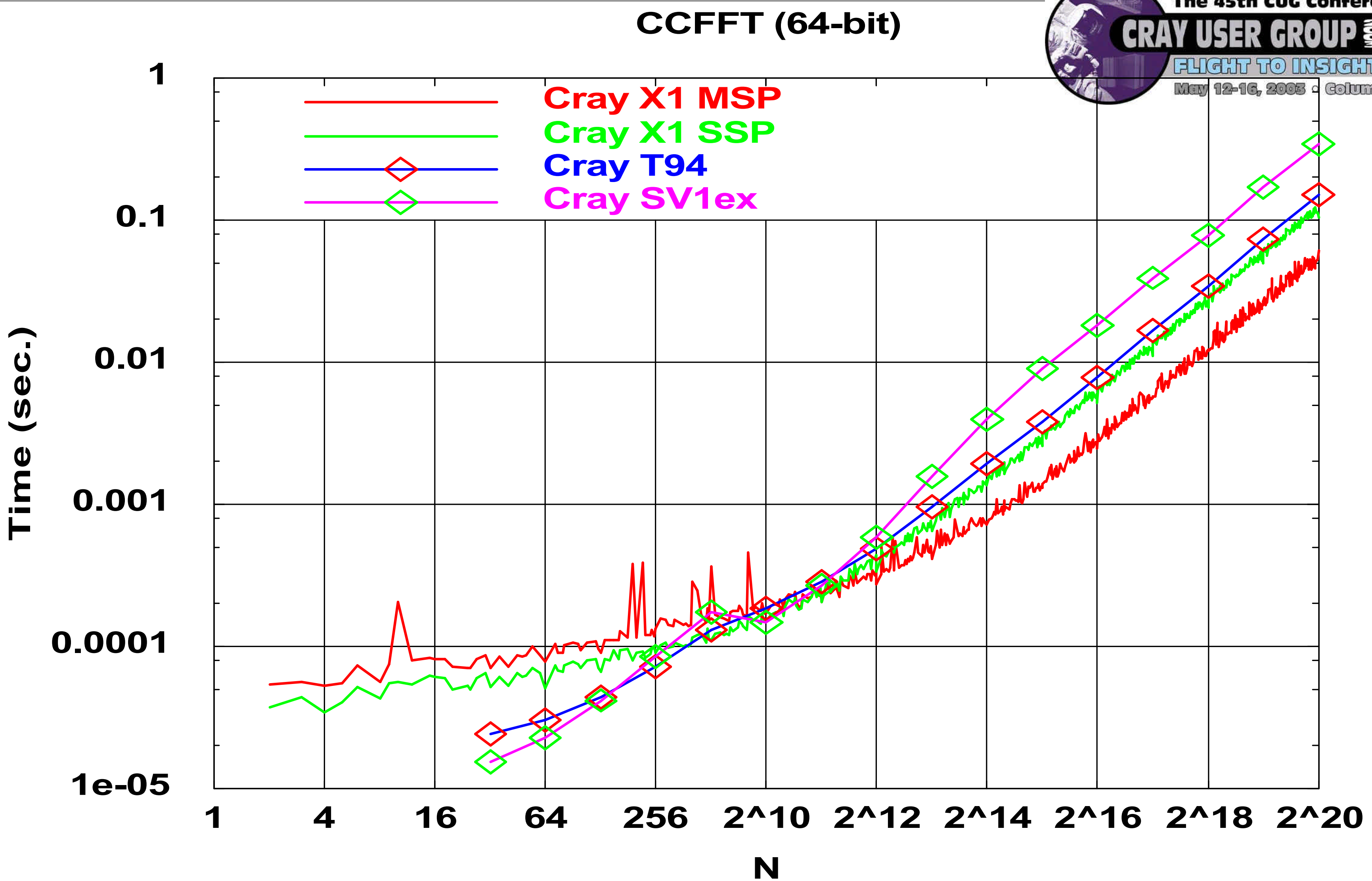
The 45th CUG Conference
CRAY USER GROUP
FLIGHT TO INSIGHT
May 12-16, 2003 ○ Columbus, Ohio
OSC



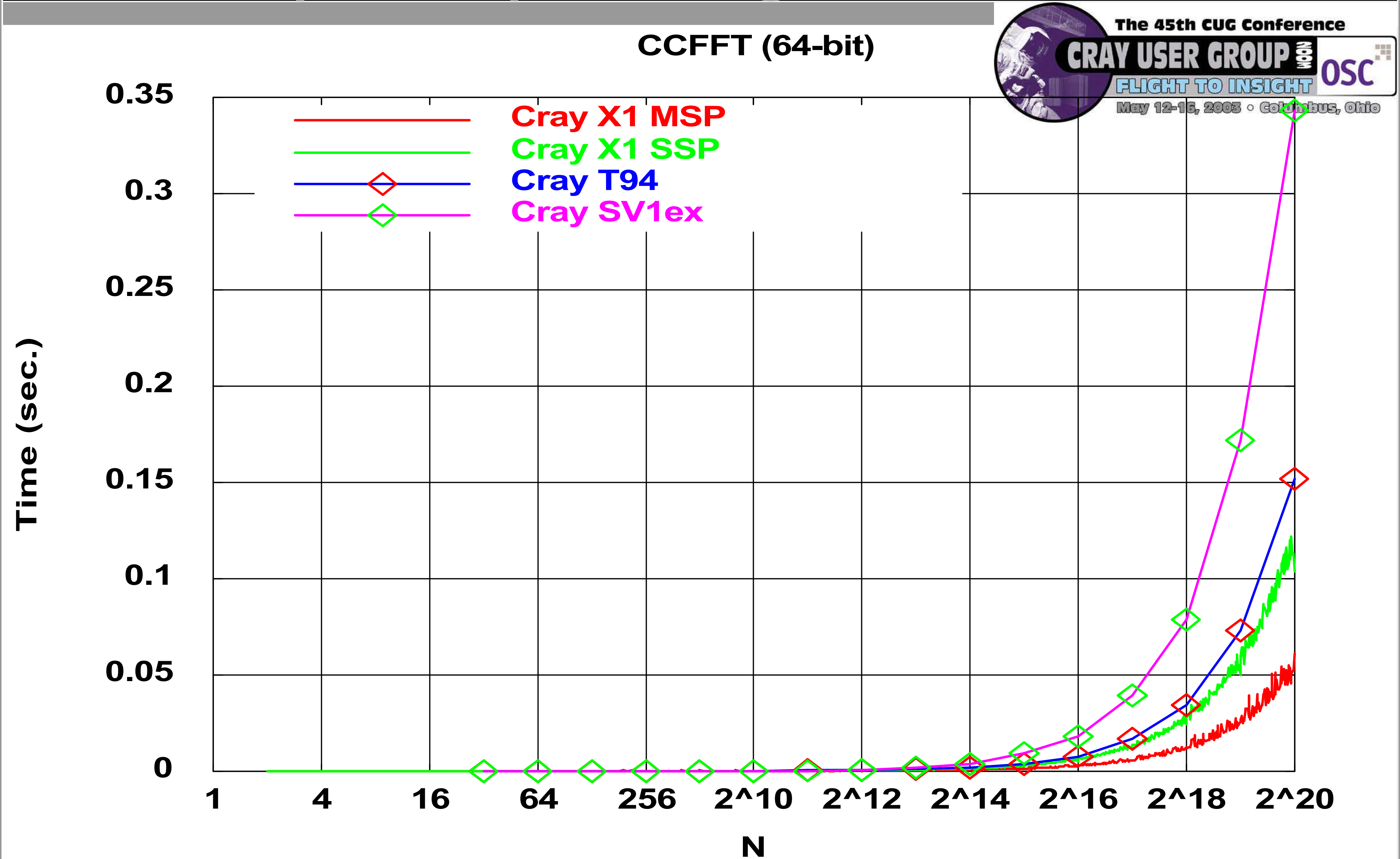
CCFFT2D (64-bit) Performance



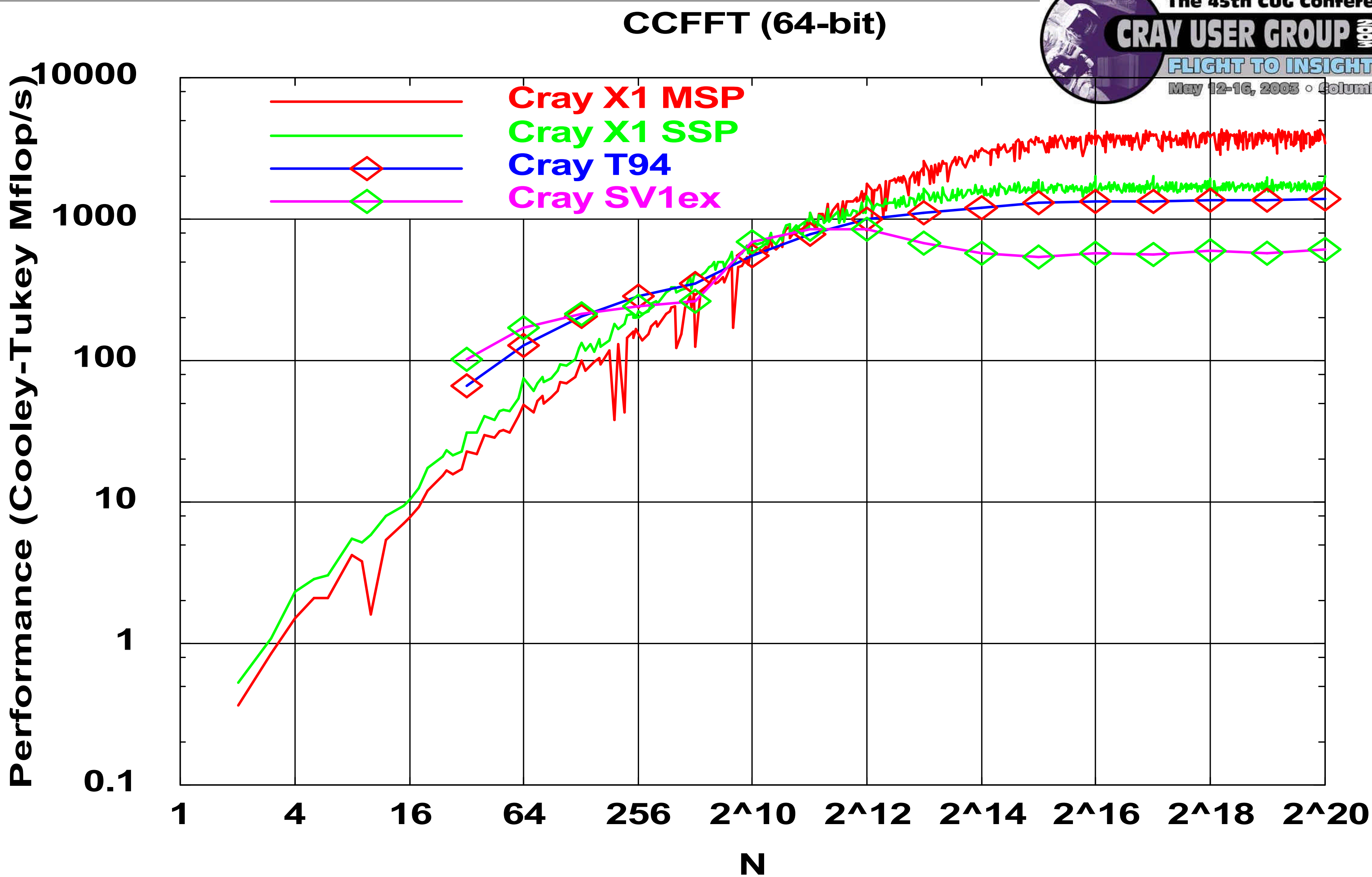
CCFFT (64-bit) Timings



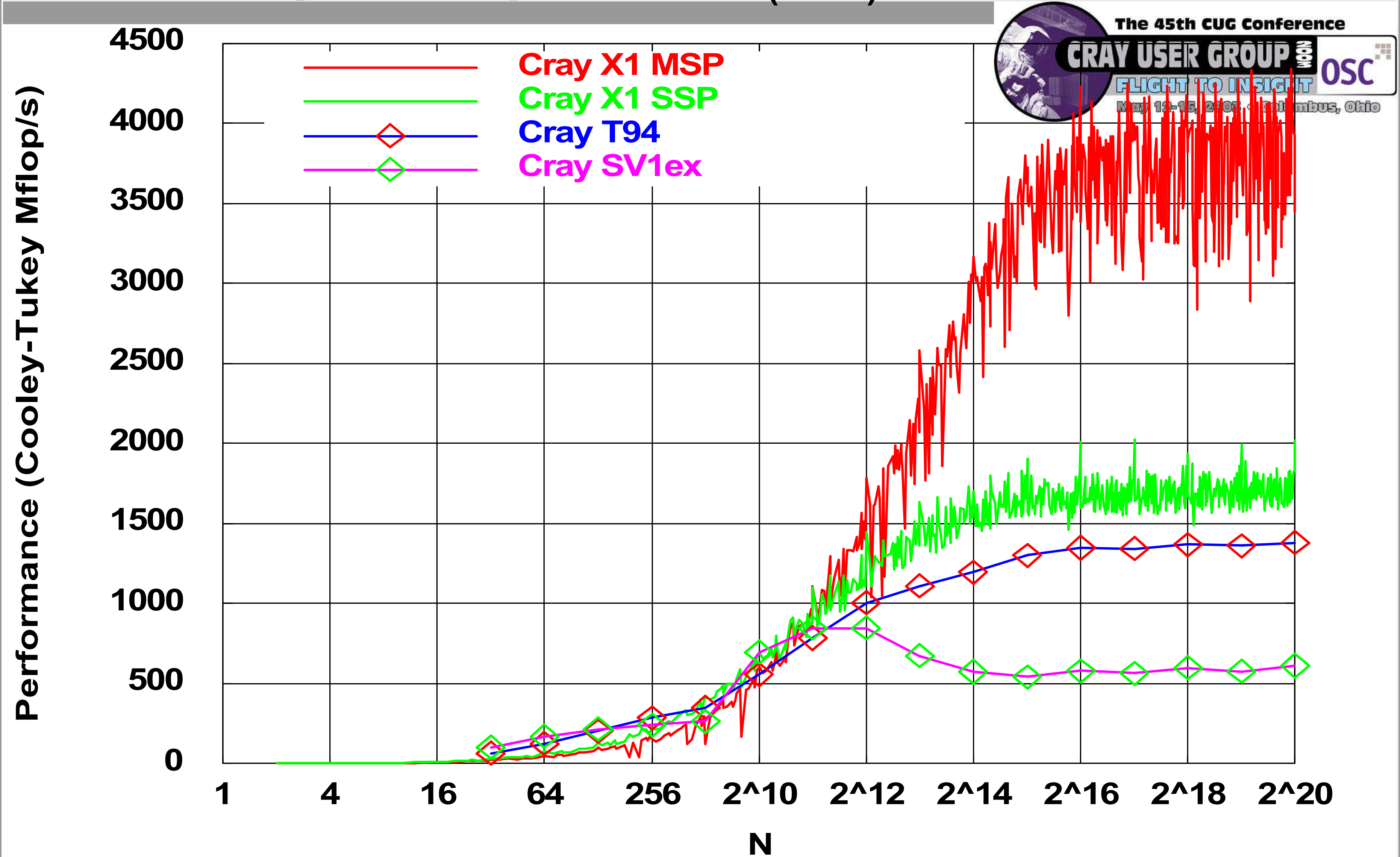
CCFFT (64-bit) Timings



CCFFT (64-bit) Performance



CCFFT (64-bit) Performance



(V) 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.