Goals of tutorial

- Introduce NMRbox platform

- Showcase NMRbox with NUS tools
  - A dozen different NUS processing tools installed and configured – more coming.

- Demonstrate potential of NMRbox
  - Now that the platform is maturing we can focus attention to **enhancing tools**, providing training materials, etc.
  - For this tutorial – Created a prototype, ***autonus***, for wrapping NUS tools into a single package.

- Learn some information about NUS
  - NUS can improve:
    - Time – if there is sensitivity to spare
    - Resolution – Collect to longer time increments without increase experiment time
    - Sensitivity per unit time – Focus more sampling to increments with smaller time increments without sacrificing resolution
  - Learn about features / characteristics of different techniques
  - Example scripts for various NUS techniques

NUS Processing Tools

Tool	Data Directory after Processing
dft-nmrpipe	dft-nmrpipe
dft-rnmrtk	rnmrtk
MaxEnt (RNMRTK)	maxent
LONER (RNMRTK)	loner
hmsIST	hmsist
NMRPipe IST	nmrpipe-ist
NESTA-NMR	nestanmr
SMILE	smile
NMRFx-Processor ISTMATRIX	The directory where you start nmrfxp
NMRFx-Processor NESTA	The directory where you start nmrfxp
CAMERA (MaxEnt)	camera
nmr_wash (SCRUB)	scrub
nmr_wash (CLEAN)	clean

Differences in NUS techniques

- Acquisition dimension (DFT)
- **maxent & loner**: Output frequency – no DFT of indirect dimensions
- **scrub & clean**: Start with nuDFT spectrum and scrub sampling artifacts
- **hmsIST, NMRPipe-IST, CAMERA, NESTA-NMR, SMILE, NMRFX-Processor IST & NESTA, MDD & MDD-CS**,:"fill-in" missing time domain data – data processed with DFT afterwards
- Keep or replace experimental data?
  - Keeping experimental data is akin to overfitting the data
- Phases: Some tools are independent of phase, some need to know the phase during reconstruction, and some will not work properly with a first order phase correction.
- Extending data beyond last collected point
- Deconvolution: MaxEnt can deconvolve linewidth and J-couplings
- Non-linearity

HNCACB 12 kDa four helix bundle

Experimental setup

- grpdly = 67.9862060546875

	t1	t2	t3
Nucleus	N	CACB	HN
Echo-Antiecho	no	no	no
Reference	119.087	47.742	4.773
Phases	0, 0	0, 0	90, 0
ZF size	256	256	1024
FT options	alt, neg	alt	
Extract Region	9.52 – 5.9 ppm		
Sampling	800 points (hyper-complex) (21%)		1024 (complex)
Max increment	50	76	1024

HNCACB Sample Schedule (nuslist)

"cd hncacb_nus"
"more nuslist"
"cat nuslist | sort –n -k2 -k1 | more"

- Note: Data is hyper-complex in all three dimensions.
- Note: Sample schedule (zero-indexed). Processing parameter files (one-indexed)

Sample schedule		**Pts (C, Total)**	**Compressed**	**Expanded**	**Expanded**
t1	**t2**	**x (t3)(HN)**	**FID #**	**y (t1)(N)**	**z (t2)(C)**
0	0	1024, 2048	1-4	1, 2	1, 2
2	0	1024, 2048	5-8	5, 6	1, 2
5	0	1024, 2048	9-12	11, 12	1, 2
12	0	1024, 2048	13-16	25, 26	1, 2
15	0	1024, 2048	17-20	31, 32	1, 2
…	…	…		…	…
20	73	1024, 2048		41, 42	147, 148
31	73	1024, 2048		63, 64	147, 148
36	73	1024, 2048		73, 74	147, 148
	74 (missing)				149, 150
49	75	1024, 2048		99, 100	151, 152

Total number of nmrPipe planes

Expanding NUS data

We will see if a bit how we expand NUS data, but lets take a look at some NUS data that is already expanded.

"cd HNCACB_nus/data"
"nmrDraw -in test%03d.ft3"

For z (t2) planes 1 and 2 we see identical FIDs at y (t1) positions of 1,2,5,6,11,12, etc.
"Goto z plane 147 and FIDs for y (t1) positions of 41, 42, 63, 64, 73, and 74

Thus, The NUS data has been fully expanded with zeros for any FIDs not in the sample schedule and the collected time domain data for points in the sample schedule.

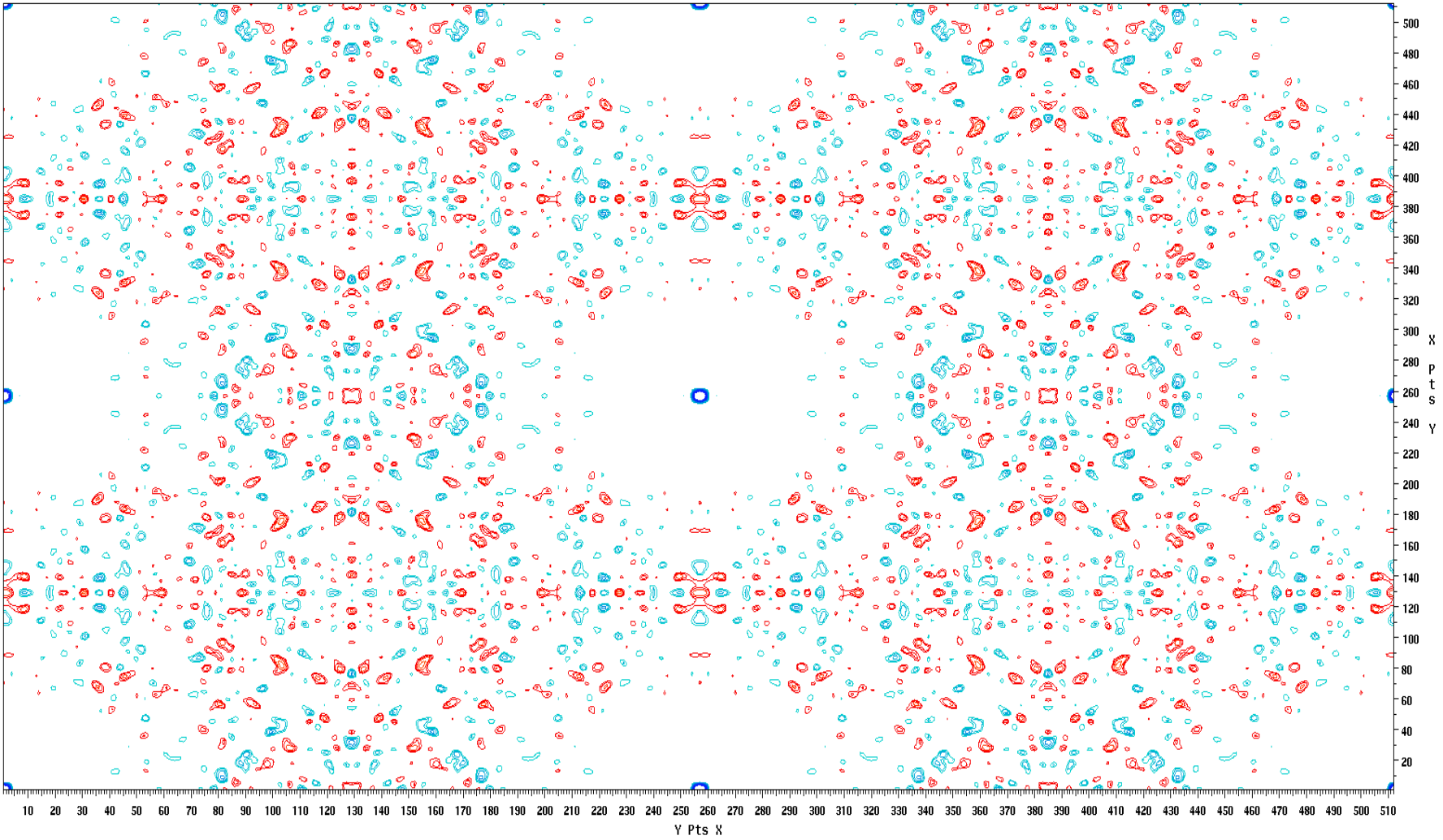
Benefits:

- Perform fixes for sensitivity enhancement during conversion – leads to consistency in processing parameters
- Allows a nuDFT for a quick view of the data.

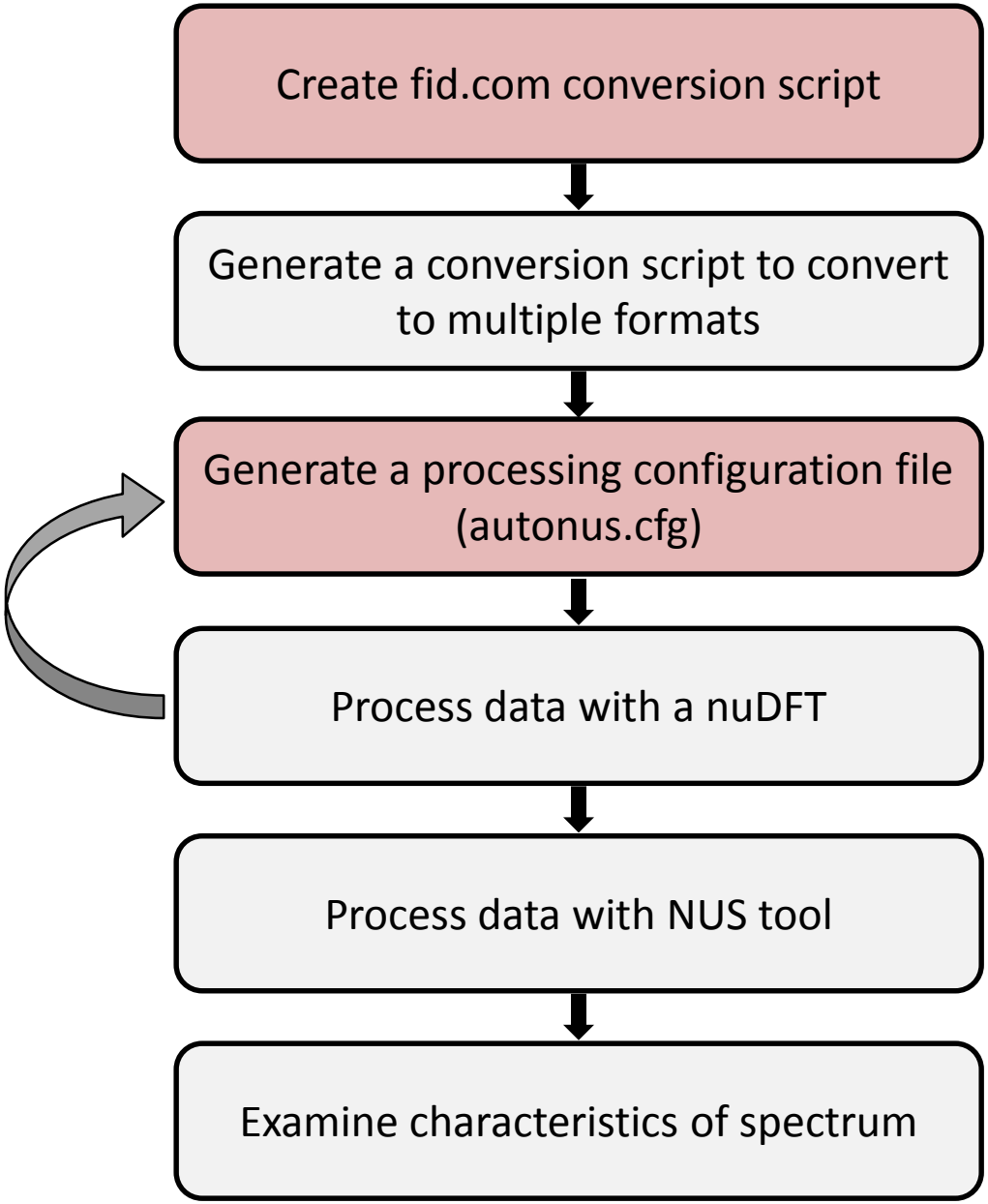
Expanded y (t1)(N)	Expanded z (t2)(C)
1, 2	1, 2
5, 6	1, 2
11, 12	1, 2
25, 26	1, 2
31, 32	1, 2
…	…
41, 42	147, 148
63, 64	147, 148
73, 74	147, 148
	149, 150
99, 100	151, 152

HNCACB Sample Schedule Point Spread Function

By applying a DFT of the sample schedule with a 1 for values in the sample schedule and 0 for missing values you obtain a point spread function (PSF). The PSF is convolved with all signals in the spectrum leading to “sampling noise” in the spectrum.



autonus - General Workflow



nuDFT File conversion (fid.com)

“cd HNCACB_nus”
“bruker”

- Press “Read Parameters”; Change output template from ”./ fid/” to “./data/”
- Press “Save Script”
- Press “Quit”

“more fid.com”

```
#!/bin/csh

nusExpand.tcl -mode bruker -sampleCount 800 -off 0 \
  -in ./ser -out ./ser_full -sample ./nuslist

bruk2pipe -in ./ser full \
  -bad 0.0 -ext -aswap -AMX -decim 2040 -dspfvs 20 -grpdly 67.9862060546875  \
  -xN              2048  -yN               100  -zN               152  \
  -xT              1024  -yT                50  -zT                76  \
  -xMODE            DQD  -yMODE        Complex  -zMODE        Complex  \
  -xSW         9803.922  -ySW         2413.127  -zSW        12315.271  \
  -xOBS         700.473  -yOBS          70.986  -zOBS         176.141  \
  -xCAR           4.773  -yCAR         119.087  -zCAR          47.742  \
  -xLAB              HN  -yLAB             15N  -zLAB             13C  \
  -ndim               3  -aq2D         Complex                        \
  -out ./fid/test%03d.fid -verb -ov

xyz2pipe -in ./fid/test%03d.fid \
| nusExpand.tcl -mask -noexpand -mode pipe -sampleCount 800 -off 0 \
  -in stdin -out ./mask/test%03d.fid -sample ./nuslist

sleep 5
```

Generate processing configuration file (autonus.cfg)

For the workshop the autonus.cfg files have been pre-created for speed.

The file format is ugly and cumbersome

- Need to decide if autonus will be a separate GUI based program or whether to wrap it into NMRFx-Processor

“more autonus.cfg”

- The text file contains some basic input questions and selections for processing along the acquisition dimension, t1, t2, as well as sections for the NUS tools.

- Initially one can guess as to some of the values, but the workflow is to process data with a nuDFT to determine values such as phases, sign alterations, apodization functions, etc. This will likely need to be done in an iterative process.

- For the workshop the values should be set appropriately for the HNCACB spectrum without any editing necessary

- Restore fid.com and autonus.cfg to defaults for Scripps workshop if necessary.
  - “cp fid.com.scripps fid.com”
  - “cp autonus.cfg.scripps autonus.cfg”

Create conversion scripts

Before any data processing can be performed the data must be converted to the correct format.

- All the NUS tools utilize the nmrPipe format except rnmrtk, maxent, and loner.
- The autonus tool will attempt to create a single conversion script that will read in Varian or Bruker data and convert to all the formats.
- Note that while a bit slow, the step only needs to be performed once.

“autonus convert”

“more convert.com”

- All the conversion scripts concatenated. Individual scripts are also created.

“./convert.com”

- This will run all the conversion scripts
- All data saved in current directory under data

nmrpipe & rnmrtk formats are fully expanded with zeros in place of any missing FIDs.

- Allows a nuDFT of the spectrum

nuDFT with nmrpipe

“cd HNCACB_nus”
“autonus dft-nmrpipe”
“ls -ltr”
“more dft-nmrpipe.com”

Script broken into 6 sections

1. Acquisition dimension for expanded data (x saved back to x [xyz])
    - Intermediate files will be used by:
        - dft-nmrpipe, nmrpipe-ist, clean, & scrub
2. Acquisition dimension for expanded data for nestanmr (x saved back to x [xyz])
    - Intermediate files will be used by:
        - nestanmr
3. Acquisition dimension for expanded data (x saved to z [yzx])
    - Intermediate files will be used by:
        - smile
4. Acquisition dimension for compressed data (x saved to z [yzx])
    - Intermediate files will be used by:
        - Hmsist & camera
5. Process t1 dimension for nuDFT (y saved back to y)
6. Process t2 dimension for nuDFT (z saved to y)

nuDFT (nmrpipe)

“./dft-nmrpipe.com”
“./ls -ltr”

- nuDFT data saved in dft-nmrpipe directory.
- Intermediate t3 processed data saved in dft-nmrpipe, xyz and yzx folders for later use

“nmrDraw -in dft-nmrpipe/dft%03d.ft3 “

Move through various planes and examine the large amount of “noise”. The noise is both the empirical noise and “sampling noise” due to the convolution of the point-spread-function with every signal in the spectrum

The NUS processing tools will attempt to remove the “sampling noise” from the final spectrum.

nuDFT (rnmrtk)

“cd HNCACB_nus”
“autonus dft-rnmrtk”
“more dft-rnmrtk.com”

- Script processes data in all three dimensions with nuDFT

- Two intermediate files are generated
  - noisecalc.sec
    - Used by noisecalc to estimate values for DEF and AIM for MaxEnt reconstruction
  - f3_proc.sec
    - Used by MaxEnt and LONER as starting point for reconstruction in the indirect dimensions

“nmrDraw -in rnmrtk/dft-rnmrtk_f3f2.ft3”
- Intermediate files ready for automatic MaxEnt reconstruction

Auto MaxEnt Reconstruction

The maximum entropy algorithm in RNMRTK has three adjustable parameters; **def**, **aim**, and **lambda**. Data can be processed with ***constant aim*** mode or ***constant lambda*** mode.

Workflow for 3D processing with msa2d

Choose Def and Aim (Based on empirical noise)

Process spectrum in Constant Aim mode

Look up the converged Lambda value

Process the whole dataset in Constant Lambda mode

Choosing def and aim

- ***aim*** should be a small integer multiple of the rms noise in the data
  - compute the rms for a blank region of a 1D (typically use fid with weakest signal)
- ***def*** should be a value smaller than that of the smallest expected peak
  - too small and the baseline noise distribution becomes “spikey”
  - too large and the noise level will be too high
- ***noisecalc*** – A tool developed by Mehdi Mobli to compute reasonable values of *def* and *aim* given a blank region of a 1D spectrum
- ***autonus*** – A tool that will perform all these steps in an automated fashion

Auto MaxEnt Reconstruction

“cd HNCACB_nus”

“autonus maxent”

- Script performs the Workflow steps of the previous slide automatically
  - Reads in t3 processed data
  - Analyzes spectrum for noise and estimates DEF and AIM
  - Processes whole spectrum in constant AIM mode
  - Examines the log file and averages converged Lambda values
  - Processes whole spectrum in constant Lambda mode
  - Saves data and outputs some information
    - Maximum number of loops
      - Values between 20 and 80 are typical
    - Reports DEF, AIM, and LAMBDA

“ls -ltr”

- maxent.com (Processing script)
- msa2d_param (MaxEnt paramter file)
- msa2d.txt (MaxEnt reconstruction log)

Auto MaxEnt Reconstruction

"more msa2d_test_param" "more msa2d_param"

```
DEBUG 2
NLOOPS 1000
DEF 3.9189000
LAMBDA 0.09208
SCALEFIRST 0.5
SCHED ./tlkt.scd
NOUT 256 256
PHASE -0.0 0.0 0.0 0.0
LW 0.0 0.0
JVALUE 0.0 0.0
```

DEF determined from noisecalc multiplied by "def multiplication factor"

LAMBDA determined from trial run in constant AIM mode

Final data size extrapolated to 256x256

To re-run with different parameters, if desired, simply edit the msa2d_param file and execute "./maxent.com"

"nmrDraw -in maxent/maxent_f3f2.ft3"

Automatic LONER Reconstruction

The l_1-norm real algorithm in RNMRTK has one adjustable parameters; **aim**

Processing Workflow

- Read in t3 processed data
- Analyzes spectrum for noise and estimates AIM
- Reconstruct indirect dimensions with LONER
- Saves data and outputs some information
  - Maximum number of loops
  - Reports AIM

“more loner_param”

```
DEBUG 2
NLOOPS 1000
AIM 354.6980000
SCHED ./tlkt.scd
NOUT 256 256
LPHASE -0.0 0.0
LW 0.0 0.0
JVALUE 0.0 0.0
```

“nmrDraw –in loner/loner_f3f2.ft3”

hmsIST, nmrpipe-ist, SMILE, NESTA-NMR, CAMERA, NMRFx-IST, & NMRFx-NESTA Workflows

Workflow

Convert time domain data

Process the acquisition dimension with NMRPipe

Reconstruct missing time domain data

Process two indirect dimensions with NMRPipe

NMRFx-Processor (nmrfxp)

NMRFx-Processor is a relatively new NMR Processing package from Bruce Johnson at CUNY.

Fully GUI based program

Lots of intelligence built-in

Handles Bruker, Varian, and other data

Processing scripts are built in the GUI, but the executed script (process.py by default) is a python script which can be saved and run in a stand-alone manner (not yet implemented in NMRbox)

Has IST and NESTA algorithms implemented

Has the framework to allow calls to external programs from within the python scripts

Under active development

“cd HNCACB_nus”
“nmrfxp”

NMRFx-Processor (nmrfxp) Instructions

File --> Open and Draw. Navigate to Scripps-2017/HNCACB_nus and select ser and open
From the Dialog box without the spectrum choose Scripts and Auto Generate
For ref select ppm at center and H20
Select the Operations Tab and make sure D1 dimension is selected
    Add TD-Solvent TDSS
    Select ZF and add 1024 for size
    Select phase and change ph0 to 90 and select dimag
    Add Regions --> EXTRACT and set start=160 and end=415
Select D2,3 then select ISTMATRIX and check the disabled checkbox
Select D2
    Change ZF size to 256
    Change FT to turn off negateImag
    Select the dimag checkbox for phase
Select D3
    Change ZF size to 256
    Change phase to -23.0 50.0 and select the dimag checkbox
Click the process button at the bottom – performs a nuDFT
After processing select Y and choose 13C
Select Z and Center – then move about the spectrum
Go back to D2,3
    Turn off the disable button & click Process
Disable ISTMATRIX
    Add Sampling --> NESTA & ReProcess

hmsIST

“autonus hmsist”
“ls -ltr”

- hmsist.com (IST script)
- phf2pipe.com (Script to rearrange data)
- run-hmsist.com (Script to run hmsist.com and parallalize calculations)
- hmsist-ft23.com (Script to process t1/t2 dimensions after hmsIST_
- hmsist-all.com (Script to run them all)

“more hmsist.com”
“hmsIST –help”

```
hmsIST -dim 2 -incr 1 -xN 128 -yN 128 -user 1  \
   -itr 800 -verb 0 -ref 0 -sched ./nuslist \
   < ./yzx/${in} >! ./yzx_ist/${out}
```

“./hmisist-all.com”
“nmrDraw –in hmsist/hmsist%03d.ft3”

NMRPipe IST

"autonus nmrpipe-ist"
"more nmrpipe-ist.com"

Notes

- ist3D.com processes the acquisition dimension with FT, performs an IST calculation to fill in missing points in t1/t2, and then processes the t1/t2 planes with FT all internally.
- Arguments can be passed to change default values for phases, FT arguments, apodization, etc.
- In the latest version -istMaxRes can be set to Auto
- Also in latest version data can be extended beyond the last point collected – slows calculation
- Likely the slowest of all the NUS techniques – especially when the highest increment values are large.

"./nmrpipe-ist.com"
"nmrDraw -in nmrpipe-ist/ist%03d.ft3"

NESTA-NMR

"cd hncacb_nus"
"autonus nestanmr"
"ls -ltr"
"more nestanmr.com"

```
# Notes on parameters
# --fids    =  Location of FIDs with the acquisition dimension processed
# --nuslist =  NUS sample schedule (zero indexed)
# --outdir  =  Location where output is saved. Note that files will be saved with a .ft1 extension
# --outname =  Base name of the files to be saved
# --method  =  Regularization method (1 - L1/IRL1) (2 - Gaussian-SL0)
# --iter    =  Maximum number of interations (10-50 for L1/IRL1) (>= 5000 for Gaussian-SL0)
# --rwiter  =  Number of re-weighted iterations (1 = L1) (2-10 = IRL1). Computation time goes up with rwiter.


nestanmr \
--fids ./nesta/test%03d.dat \
--nuslist ./nuslist \
--outdir ./nesta \
--outname test \
--method 1 \
--iter 50 \
--rwiter 10 \
```

"./nestanmr.com"
"nmrDraw -in nestanmr/nestanmr%03d.ft3"

SMILE, CAMERA, SCRUB

“cd HNCACB_nus”
“autonus smile”
“ls -ltr”
“more smile.com”
“./smile.com”
“nmrDraw –in smile/smile%03d.ft3”

Repeat for camera and scrub

Additional Data Sets

- 15N-NOESYHSQC.fid – A Varian data set collected uniformly and sub-sampled into a NUS with 26% coverage.
- w
- Process with autonus to see how the NUS tools work with a high dynamic range spectrum
  - Note: Don’t use NMRPipe-IST

Uniform HNCO (ni=128, ni2=128) with MaxEnt

“cd hnco-large.fid”

- The HNCO was run with t1 and t2 collected out to 128 points.
- The process.com script processes the data with a sample schedule entered as an argument with RNMRTK’s maximum entropy reconstruction and performs a nuDFT. Any point not in the sample schedule will be ignored during processing.
- The only criteria is that the sample schedule not exceed 128 in either dimension and that the schedule is 1 indexed.
- The output files are:
  - BaseOutputName_nudft.ft3 (3D) & BaseOutputName_nudft_proj.ft2 (Projection)
  - BaseOutputName_msa2d.ft3 (3D) & BaseOutputName_msa2d_proj.ft2 (Projection)

“more rnmrtk.com”

- Script to process HNCO with FT in all dimensions with RNMRTK

“more nmrpipe.com”

- Script to process HNCO with FT in all dimensions with NMRPipe

“./rnmrtk.com”

- Create a sample schedule making sure that increment values do not exceed 128 in either of the two dimensions

“./msa2d.com SampleSchedule.scd BaseOutputName”

- After running the data will be saved with the filename Base_output_name.ft3