MIPS-24㎛m Pipelines Review (S6.3 Delivery)

Frank Masci

MIPS-24 Cognizant Engineer

November 6, 2002

MIPS-24 Pipeline Summary

? In all, there are twelve automated pipeline threads for processing raw science Data Collection Events (DCEs) to produce Basic Calibrated Data images (BCDs) containing actual pointing and formatted headers:

1. SUR-mode science (2-plane DCEs: slope + difference)
2. RAW-mode science (multi-plane DCEs: Either 6, 8, 20 or 60 plane data ramps)
3. SUR-mode dark-current calibration (pre-processing)
4. SUR-mode dark-current calibration (ensemble-proc of 3)
5. RAW-mode dark-current calibration (pre-processing)
6. RAW-mode dark-current calibration (ensemble-proc of 5)
7. Non-linearity calibration (pre-processing)
8. Non-linearity calibration (ensemble-proc. of 7)
9. Flat-Field (non-uniformity) calibration (pre-processing)
10. Flat-Field (non-uniformity) calibration (ensemble of 9)
11. Boresight ? FPA pointing transfer thread and Final Product Generator (FPG).
12. Pipe0 or “zip” science pipeline (for system validation and quick generation of raw translated headers with pointing.

? Threads **4, 6, 8** and **10** produce calibration products required by science threads 1 and 2.

SUR-mode Science Pipeline Thread

INPUT SUR-Mode 2-plane Science DCE

↓

SANITY_DATATYPE – FITS header keywords sanity check
- Instrument/Channel/Mode/Data Missing

↓

QATOOL_DCE – Quality assurance characterization/Statistics on DCE

↓

SANITY_CHECK – Image-statistics sanity check from QATOOL_DCE output

↓

TRANHEAD – Translation of FOS keywords, derivation of exposure-time keywords, insertion of required keywords

↓

CALTRANS – Acquire/transfer desired calibration data from database

↓

CVTI2R4 – Conversion from I*2 to R*4 & 0.5 DN truncation correction
- Creates D-Mask and replaces missing data with NaNs
- Flags pixels which are hard saturated after A-to-D

↓

SATMASK – Saturated-pixel detection: Output to D-Mask

↓

CVTDNSEC – Conversion from DN/sampling-time to units of DN/seconds

↓

SLOPERROR – Estimate noise in slope image

↓

DESATSLOPE – Slope desaturation for robust droop correction

↓

DROOPOP – Droop correction

↓

ROWDROOP – Row-droop (mux-bleed) correction

↓

PTO

SUR-mode Science Pipeline Contd..

CUBESUB_SUR – Dark-current subtraction from slope image

SLOPECORR – Non-linearity correction (slope data only)

FLATAP – Flat-field, non-uniformity correction

CAL

RADHITMEDFILT – Single frame rad-hit detection

DNTOFLUX – Conversion from DN/sec to μJy/arcsec^2

REPLACEPIXELS – Replace saturated (slope) pixels with pixels from difference image, output to new single plane slope image (replaced.fits)

SPLITSURIMAGE – Split 2-plane science and uncertainty image cubes into slope and difference images. Remove second plane from d-mask.

QATOOL_DP - Quality assurance characterization/Statistics on BCD

CHECKMASK_DP – Statistics on D/B-Mask of slope image

QALOADER_DP – Load output from QATOOL_DP into database

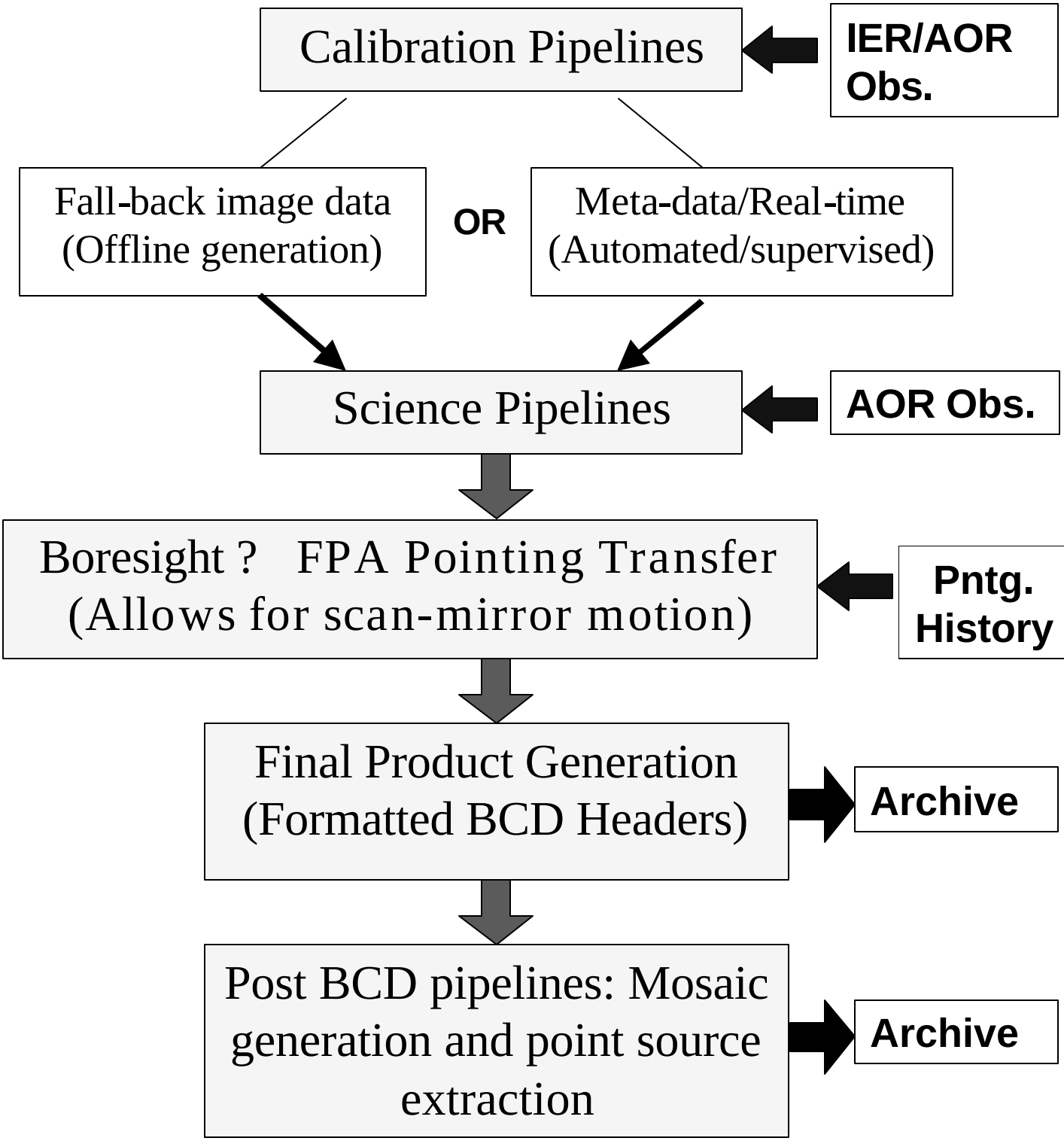
OUTPUT BCD (Basic Calibrated Data) FITS images (slope, difference, replaced, uncertainty and B-Mask images) ready for archiving.

Pipeline Parameters

- Defined in “namelist” configuration files for each pipeline module and will be loaded by pipeline operator using The File Server (TFS).
- Currently, have 21 tunable parameters for SUR-science thread:

```
CALTRANS_MIPS1SUR_405
  CalRule = '0;0;0;0',
CUBESUBIN
  Operation = 1,
  PMASK_FATAL = 4352,
  DMASK_FATAL = 16384,
  CMASK_FATAL = 6144,
DESATSLOPEIN
  CmdFrm_Keyword = 'DCE_FRMS',
  PMaskFatal = 8192,
  DMaskFatal = 16384,
  CMaskFatal = 256,
  DMaskDESAT = 16,
DNTOFLUXIN
  ConversionFactor = 0.36,
  BunitString = microJy/arcsec^(2),
  PMaskThreshold = 16384,
  DMaskThreshold = 8192,
  Data_Plane = 1,
DROOPOPIN
  Droop_Coef = 0.33,
  Droop_Error = 0.01,
  PMASK_FATAL = 16384,
  DMASK_FATAL = 16384,
  CMASK_FATAL = 16384,
FLATAPIN
  PmaskMask=4352,
  CmaskMask=16384,
  DmaskMask=8192,
QATOOLIN
  Operation = 1,
  Transpose = 1,
  NBinsHistogram= 41,
  NDevHistHalfWidth = 5,
  Data_Plane = 1,
  Box_X_Position = 1,
  Box_Y_Position = 1,
  Box_X_Width = 128,
  Box_Y_Height = 128,
  CentralFraction1 = 0.81,
  CentralFraction2 = 0.72,
  NEdge = 0,
  KeywordPrefix = 'bcd',
  StatisticsGroup = 0,
RADHITMEDFILTIN
  Threshold = 1000,
  Window = 3,
  Data_Plane = 1,
  Operation = 2,
  PMaskFatal = 16384,
  DMaskFatal = 8192,
SATMASKIN (replacepixel.nl)
  PMaskFatal = 4352,
  DMaskFatal = 16384,
  Replace_With_NaN = 1,
ROWDROOPIN
  CC_1 = 7.3e-5,
  CC_2 = 7.3e-5,
  CC_3 = 7.3e-5,
  CC_4 = 7.3e-5,
  UCC_1 = 0.5e-5,
  UCC_2 = 0.5e-5,
  UCC_3 = 0.5e-5,
  UCC_4 = 0.5e-5,
  PMASK_FATAL = 16384,
  DMASK_FATAL = 16384,
SANITY_DATATYPE_CONDITIONS
|  status            | criteria
|
 normal         1&&2&&3&&4&&11&&12&&13&&14&&16
 correct_pipeline   5||8||9
 SUR_mode_data      11&&14
 no_missing_data    17
 no_missing_groups  !18
SANITY_DATATYPE_CRITERIA
| index | keyword    | test  | value  |
  1       INSTRUME      =       MIPS
  2       CHNLNUM       =       1
  3       DCENUM        >=      0
  4       EXPID         >=      0
  5       PIPENUM       =       401
  6       PIPENUM       =       402
  7       PIPENUM       =       403
  8       PIPENUM       =       404
  9       PIPENUM       =       405
 10       PIPENUM       =       406
 11       NAXIS         =       3
 12       NAXIS1        =       128
 13       NAXIS2        =       128
 14       NAXIS3        =       2
 15       NAXIS3        >       2
 16       GROUPS        >       1
 17       MISSDATA      =       F
 18       MANCPKT       >       0
SATMASKIN
  Diff_Sat_Threshold = 1000,
  PMaskFatal = 4352,
  DMaskFatal = 16384,
  DMask_sat = 8192,
  Replace_With_NaN = 1,
SLOPECORRIN
  PMaskFatal = 4352,
  DMaskFatal = 8192,
  CMaskFatal = 16384,
  DMaskNotLin = 4096,
SLOPERRIN
  Log_Filename = 'stdout',
  Read_Noise = 27,
  Confusion_Sigma = 0,
  PMaskFatal = 16384,
  DMaskFatal = 16384,
TRANHEADIN
  Operation = 1,
  Key_Suffix = '02',
  CentralFraction1 = 0.9,
  CentralFraction2 = 0.9,
  Gain1 = 5.0,
  Gain2 = 5.0,
  Gain3 = 5.0,
  Gain4 = 5.0,
  Ignore_Frames_1st = 1,
  Ignore_Frames_2nd_and_higher = 1,
```

Processing Flow



Fall-back vs. Meta-data Calibration Files

- An automated science pipeline can either use fallback and/or meta-data calibration image data:

1. **Fallback:** "Perfect/favorite" calibration data known apriori to apply within a pre-specified time-range during operations (default data to use).
   - Requires human intervention: can be made using either offline pipelines (more flexibility for ISTs/IT), or, pick products made by automated pipelines.
   - Parameters which define this are loaded into the "mips1FallBack" table of the SODB using a script by the pipeline operator: Parameters:
     - Time range (required): SCLK_start ? SCLK_end
     - Temperature/Bias Voltages.
     - "Blessed" status flag also loaded in database.

2. **Meta-data:** In flight calibration data acquired contemporaneously (or close enough) with science data.
   - Generated by automated calibration pipelines with keyword parameter values loaded automatically into "mips1MetaData" SODB table.
   - Database entries must be "blessed" before using.
   - When processing science DCE, can pick "closest-in-time" calibration product or make product interpolated at science DCE-time between two calibration products.

Calibration Transfer

- In a science pipeline, calibration transfer method can be selected by specifying desired "rule" in `caltrans` namelist file:

```
CalType  = 'dark; linearity; flat; pmask',
CalRule  = ' 0;      300;    100;   200',
```

- Rule 0: pick fall-back product where science DCE satisfies

$$T_{cal}(\text{start}) < T_{science} < T_{cal}(\text{end})$$

- Rule 100: pick closest in time before: $T_{cal} < T_{science}$
- Rule 200: pick closest in time after: $T_{cal} > T_{science}$
- Rule 300: pick either closest-in-time before, or, after.
- Rule 400: "compute" calibration product by interpolating between closest in time before and after.

- Can specify "mixed" rules like in above example.
- If closest in time product does not exist, defaults to fallback calibration product with a warning issued.
- Also can specify keyword value contraints in namelist such as temperature range.
- Caltrans will only use fallback and/or meta-data products which have been blessed. (Blessed flag turned on).

Baseline Fallback Cal. Data

- Fallback calibration files (currently used) stored in calibration repository.
- Need to refine, regenerate from lab data or during mid-IOC:

```
mips24_darkest_raw1.fits
mips24_darkest_raw1_cmask.fits
mips24_darkest_raw1_uncert.fits
mips24_darkest_raw2.fits
mips24_darkest_raw2_cmask.fits
mips24_darkest_raw2_uncert.fits
mips24_darkest_sur1.fits
mips24_darkest_sur1_cmask.fits
mips24_darkest_sur1_uncert.fits
mips24_darkest_sur2.fits
mips24_darkest_sur2_cmask.fits
mips24_darkest_sur2_uncert.fits
mips24_flatfield.fits
mips24_flatfield_cmask.fits
mips24_flatfield_priorstate.fits
mips24_lincal.fits
mips24_lincal_cmask.fits
mips24_pmask.fits
mips24_PRF_Map.tbl
mips24_PRF_Map_Image.fits
mips24_PRF_mosaic.fits
mips24_raw_cdelt12_distort.tbl
mips24_sur_cdelt12_distort.tbl
```

Scheduling Calibration IERs

- Calibration pipelines process ensembles of DCEs made from Instrument Engineering Requests (IERs) according to specific rules (see below).

Scheduling Rules (For Uplink/ISTs):

ExptypeID	ReadoutMode	IER Rule
d1	SUR	**SUR-darks1:** Ensure multiple short exposures exist to attain large number DCENUM=0 DCEs. Suggest ~30-50 DCEs (TBD) each of 10 mips-second duration.
d1	SUR	**SUR-darks2:** Have one (or even two) exposures to attain large number DCENUM>0 DCEs. Suggest ~30-50 DCEs (TBD) each of 10 mips-second duration.
d1	RAW	**RAW-darks1:** Ensure multiple short exposures exist to attain large number DCENUM=0 DCEs. Suggest ~30-50 DCEs (TBD) each of 10 mips-second duration.
d1	RAW	**RAW-darks2:** Have one (or even two) exposures to attain large number DCENUM>0 DCEs. Suggest ~30-50 DCEs (TBD) each of 10 mips-second duration.
f1	SUR	**Flats:** Have one (or even two) exposures to attain large number DCENUM>0 DCEs. Suggest ~30-50 DCEs (TBD) each of 10 mips-second duration.
l1	RAW	**Linearity:** Have one (or even two) exposures to attain large number DCENUM>0 DCEs. Suggest ~30-50 DCEs (TBD) each of 10 mips-second duration.

- Ensembles will be created manually by pipeline operator for now.

Final Archived Science BCD Products

- **SUR-mode science products:**
  - **tranhead.fits** (translated MIPL header with pointing)
  - **bcd_slope.fits** (main bcd slope image with pointing)
  - **bcd_diff.fits** (slope image corresponding to above)
  - **bcd_replaced.fits** (slope image with saturated pixels replaced by values from difference images).
  - **uncert_bcd_slope.fits** (uncertainty slope image).
  - **uncert_bcd_diff.fits** (uncertainty diff. image)
  - **bmask_slope.fits** (processing status mask for slope)
  - **bmask_diff.fits** (processing status mask for diff).
  - Processing logs.
- **RAW-mode science products:**
  - **tranhead.fits** (translated MIPL header with pointing, contains all data planes in the ramp)
  - **bcd.fits** (main slope-fit image with pointing)
  - **uncert_bcd.fits** (uncertainty slope image)
  - **bmask.fits** (processing status bit-mask)
  - Processing logs.
- **"Pipe-0" science pipelines:** (for system validation and quick generation of raw-translated headers):
  - **tranhead.fits** (translated MIPL header with pointing and pixel values in 16-bit signed integers)
  - **bcd.fits** (same as tranhead.fits, but with pixel values in real/32-bit floating point precision).

Archive Contd..

- **NOTE:** Only the header-translated raw-image (tranhead.fits) and main BCD FITS image currently contain pointing/distortion keywords.

- All ancillary data products contain formatted headers as generated by the Final Product Generator.

- Automated pipelines initially write all desired products to the sandbox (temporary archive) which is accessible through the firewall by ISTs using DAFFI (Data Archive File Fetch Interface).

- The sandbox contains more information than final archived products, e.g. processing logs, calibration data used, QA diagnostic files.

- After every campaign (~7 days), all desired sandbox products will be moved to the science archive by setting a flag in the database (See appendix for current design).

B(BCD)-Masks

- A 16-bit/pixel B-Mask (BCD-Mask) is associated with each BCD science product – both SUR and RAW.
- This reports a summary of the processing steps for every pixel in the BCD.
- Currently we have 13 conditions per pixel for SUR-mode BCDs:

Bit #	Condition
0	Incomplete or questionable row-droop correction (rowdroop)
1	No row-droop correction applied (rowdroop)
2	
3	
4	Saturation corrected and slope value replaced by difference value in "replaced.fits" image (desatslope and satmask)
5	Latent-image flag (placeholder, not implemented yet)
6	Droop removed using questionable value (droopop)
7	Flat field applied using questionable value (flatap)
8	Flat field could not be applied (flatap)
9	Radhit detection (radhitmedfilt)
10	
11	Pixel masked in pmask - bad hardware state (satmask)
12	Non-linearity correction could not be computed (slopecorr)
13	Saturated (satmask)
14	Data missing in downlink (cvti2r4)
15	<reserved: sign bit>

Preliminary MIPS-24 Processing Benchmarks

- Estimated number of MIPS-24 SUR-science DCEs per PAO (nominally every 12 hours):
    - Average: 5500 (Equal mixture of 4 and 10 s exposures)
    - Maximum: 6000 (4 s exposures)

- **Benchmark**: During S6.3 segment testing, found it took ~262 minutes to process a request of 2025 DCEs using a cluster of four 400-MHz CPUs with 3 clients (jobbers) each. Includes database interaction and copying to sandbox.

  $\Rightarrow$ ~ 31.05 seconds per SUR-science DCE on 1CPU.

- Operations network consists of 24 400-MHz CPU's

- Time to process a PAO:
    - 31.05 s * 6000 DCEs / 24 CPUs =~ 129.3 minutes
- Time to process a 7-day campaign (14 PAOs per week):
    - 14 * 31.05 s * 6000 DCEs / 24 CPUs =~ 30.1 hours

- **Conclusions**: We can process $\approx$ 5.7x the data collection rate or a total throughput of ~ 66,800 DCEs/day. Can accommodate reprocessing of data every PAO.

Offline/Stand-alone pipelines

- **Offline**: A design which does not interact with the Science Operations Database (SODB), data products are not stored using the same archive structure to be used in SSC operations and, manually operated by user.

- Useful for analysis and science validation and generation of "fallback" calibration data for use in automated (lights-out) pipelines.

- Six stand-alone pipelines (two science, four calibration) written as self-contained perl scripts:
  1. SUR-Mode Science
  2. RAW-Mode Science
  3. SUR-Mode Dark Calibration (Pre+Ensemble proc.)
  4. RAW-Mode Dark Calibration (Pre+Ensemble proc.)
  5. Detector Non-linearity (Pre+Ensemble proc.)
  6. Flat-field Calibration (Pre+Ensemble proc.)

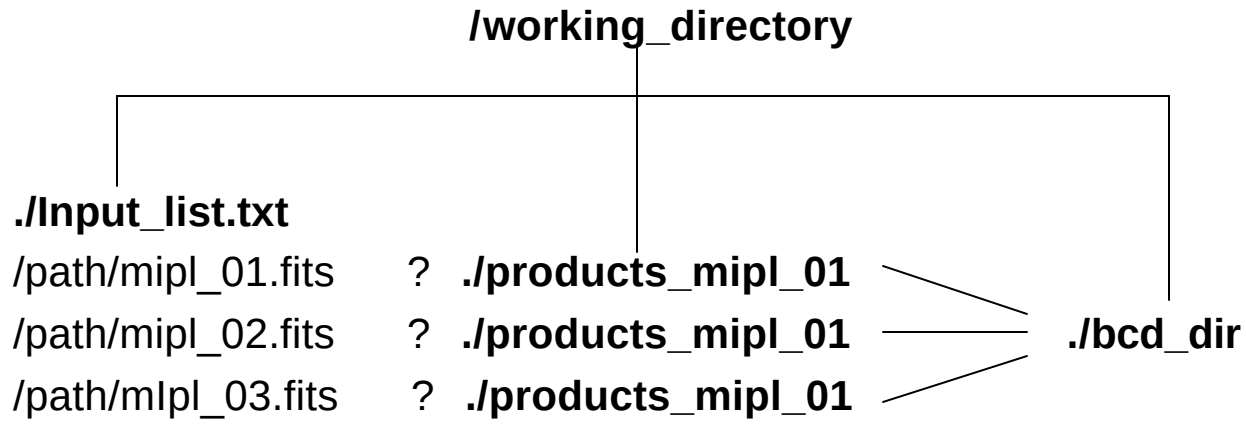
- These are easy to modify – quick turn-around for change requests.
- Example command-line execution:

```
%RUN_SURSCI.pl –i input_list.txt –d bcd_dir
```

Offline Contd..

- Namelist (parameter) files and calibration data directories are specified via environment variables.
  - ? Can have DCE dependent namelist/calibration data directories or default single directories.

- Output:
  - Intermediate products directory for each input DCE in the format products_<inp.filename>.
  - All final “BCD” products collected and placed in a single directory at the end (see example above).



- Tar file of all generically (sun-ultra/sparc platforms) built binaries and perl scripts are planned for distribution. For more information, see PDF file:

http://ssc-et.ipac.caltech.edu/Instruments/MIPS/stolovy/MIPS24_pipelines.pdf

Work to be Done (S6.3+/S7.0)

- Latent Image detection/flagging pipeline and recording information in a bit-mask.
- Record pointing keyword information in all final archived ancillary products (other than just *tranhead.fits* and main *bcd_slope.fits* images):
- Automate ensemble creation for making calibration/post-bcd products (SDM involvement).
- Retrieve essential database information/derive required keywords and write to science headers using final product generator.
- Testing, science validation, fine-tuning of namelist/threshold parameters.
- Gain map treated as calibration product (instead of read-out channel dependent gains) for DN-to-flux conversion.
- Define/refine fallback calibration data.

Appendix: Science Archive Structure

- Processed files are named using identifier numbers created in the operations database:

  Primary data product: dp<*dpId*>_<*plScriptId*>.fits

  Ancillary data products: apd<*dpId*>_<*ancilId*>.fits

- Current simplified archive directory structure: