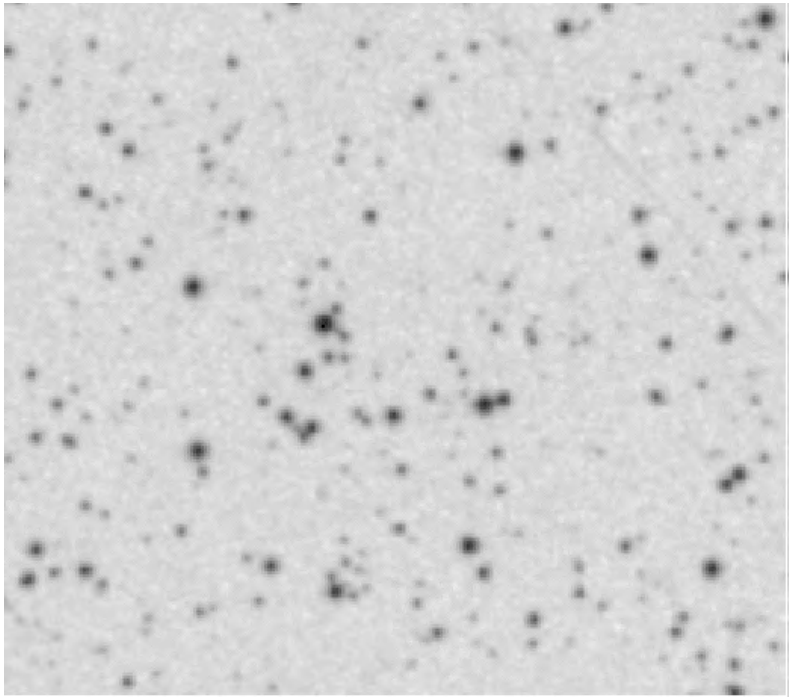
Extraction

CU6 Workshop, Royal Observatory Brussels, 12/13 October 2006

Extraction: overview

- The *Gaia*3 design selects windows on the CCD as for the astrometry and photometry
- Data from individual CCDs are available directly – no combination on board; this is helpful for cosmic rays *etc.*
- Data are collapsed spatially during the CCD readout, so no 2-dimensional information is retained or available for the ground processing (except for very bright stars V<7 where full 2-d information is available).
- Crowding is less of a problem with the *Gaia*3 design because of the larger image scale ⇒ factor ~250 less crowded – this is very helpful.
- Nevertheless, above 10000 stars/sq degree (occupation factor 1/25) crowding will remain a problem.

Simulations



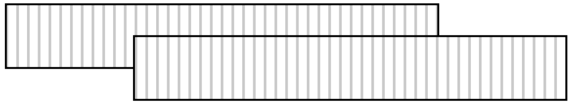
•Sky image from Palomar Sky Survey red plates, 1.75 arcmin each side centred on Galactic latitude +10.0° longitude +90.0° Star density: 15000 stars/deg^2 to V=17.0.

The simulated RVS spectra from the same region of sky on approximately the same scale:
- the limiting magnitude is V=18.5
- brightest stars in the field are V=12.5
- no readout noise in the RVS image (for clarity)
- no AC broadening introduced from the scan law.

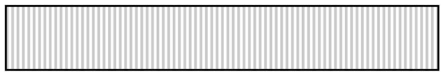
Both images on square-root intensity scales. North is up and East to the left.

Extraction: crowded windows

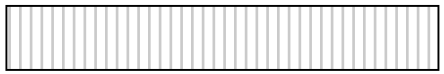
- *Gaia*3 design allows for overlapping windows:



- Also, *Gaia*3 design has two resolution modes
  - nominal resolution mode $\lambda/\Delta\Lambda \sim 11500$ for stars down to $G_{RVS} < 10$
  - low resolution mode $\lambda/\Delta\Lambda \sim 5000$ for fainter stars to reduce readout noise



or



Extraction: mixed resolution

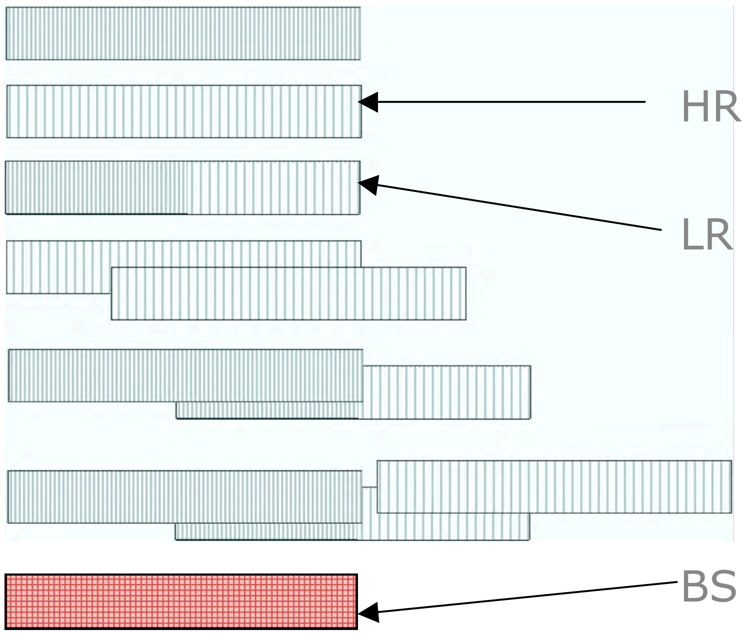
- Some spectra will have a mix of low and nominal resolution mode

as follows: 

- Faint object spectra or parts of spectra which are on the nominal resolving power sampling will be rebinned digitally on board onto the coarser sampling
- All of this leads to calibration issues (see next presentation).

RVS Window Paving

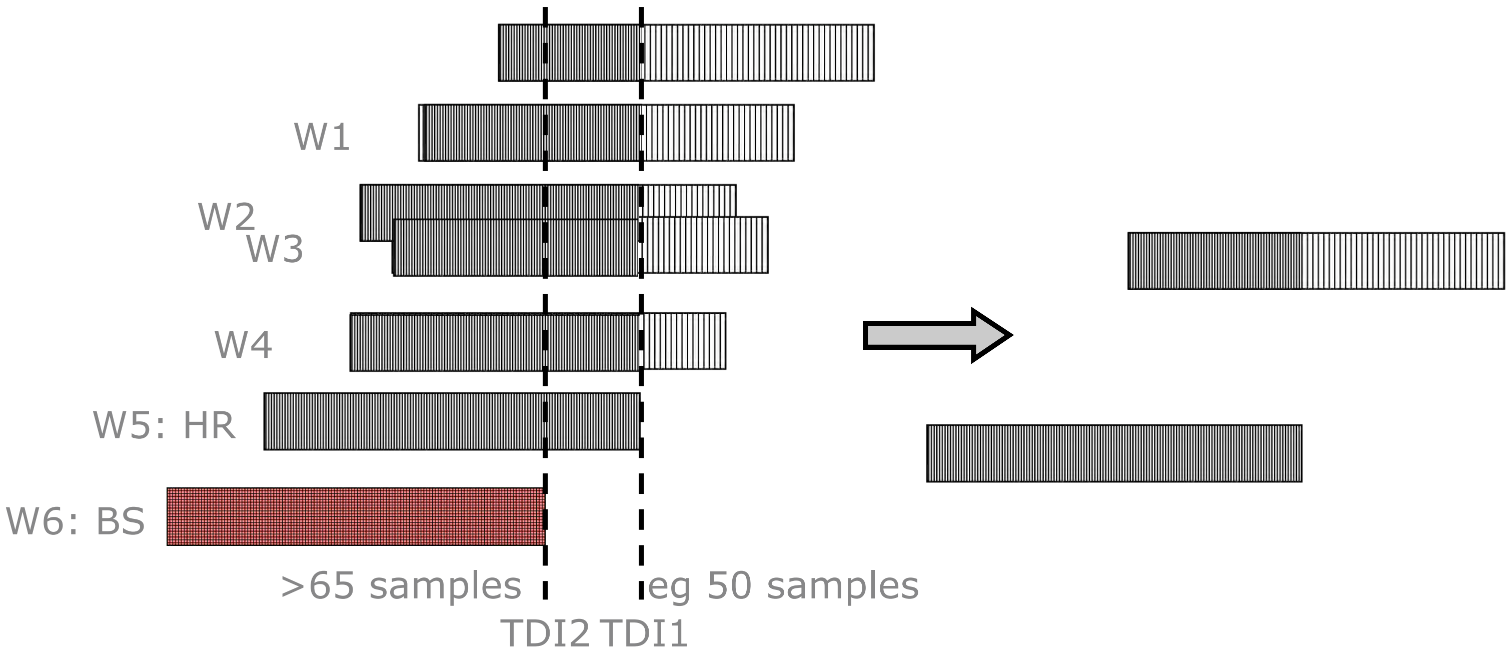
- There is a large number of possible different mixtures of windows.



Makes the extraction process complex

Priorities

- At TDI2 the bright star window is not selected for observation if there are not enough samples (<65) – implications for calibration (see later)



Extraction: modelling

- Extraction will be a complex process because of the absence of 2-d information
- In non-overlapped windows, the situation will be simple.
- For overlapped windows, much greater attention will be required to modelling the data
  - information will be needed from the RP and astrometry
  - the overlapped flux profiles will need to be predicted
  - the prediction and observed RVS fluxes will be compared
  - some iteration will be required, at later phases using the radial velocity derived from earlier scans
  - some quality of fit parameters (eg χ_ν^2) will need to be derived
- Once a good match is obtained to the observed data, the degree of mixing is known, and can be subtracted from each of the traces in the overlapped windows to produce the deblended traces.

Extraction: background

- CCD bias and any charge-injection will be removed
- The diffuse background may be available from the end of each window (or perhaps Virtual Object windows – still TBD)
- The background due to
  - point sources and
  - extended sources

  will be modelled, using information from RP and Astro.
- This information needs to be transformed backwards into instrument space from the calibrated astro and RP data

  $\Rightarrow$ 1 forward and 1 reverse calibration
- May need velocity shifts to be applied – will use past RVS measurements

Extraction: cosmic rays & calibration

- After modelling for overlapping windows, and background modelling (diffuse, point-source and extended object), spectra can be extracted and background subtracted.
- The extraction process is responsible for <u>applying</u> the calibrations
  - photometric throughput
  - wavelength scale
  - cosmetic defects on CCD

  … propagation of error/uncertainty information, quality flags
- Also, a spectrum normalised to the local continuum will be generated
- Cosmic ray removal is required using some TBD algorithm.

Multiple transits

- Such a sequence will be applied to the individual spectra from each CCD transit. Multiple transits will not be combined in the image plane as this will effectively lead to increased crowding resulting from the different orientations on the sky of the individual scans.
- There will be particular scans which for any particular object are more free from overlap than others are, and a scheme may be envisaged where the information extracted from those scans are used preferentially in the modelling necessary for those scans where spectra are overlapped.
- This may not be possible for the faintest stars. Further investigation into these possibilities is necessary.

Workpackage 620

The top level WBS is given in GAIA-C6-SP-OPM-DK-003-1

GWP-S-620 Spectra extraction

- 620-01000 Management, configuration management & interfaces
- 620-02000 Detailed functional analysis of spectra extraction
- 620-03000 Extract spectra from raw images
- 620-04000 Apply calibration
- 620-05000 Model background generated by extended sources
- 620-06000 Model background generated by point like sources
- 620-07000 Clean spectra
- 620-08000 Normalization to the continuum
- 620-09000 Validation: spectra extraction

Work Plan - Extraction

- Phase 1 (Oct06-Jan07) – mostly WP 620-01000
  - familiarisation with CU6 infrastructure
  - familiarisation with CU1 methodologies
  - set up management interfaces with CU6 and CU5 (PPARC)
  - contribute to continuing functional analysis at CU6 level
  - detailed definition of external interfaces and protocols to routines based on functional analysis
- Phase 2 (Dec06-Mar07) – mostly WP 620-02000
  - refine the requirements for the extraction workpackage tasks
  - perform detailed functional analysis for extraction tasks
  - specify internal interfaces
  - consider and devise test frameworks
- Phase 3 (Feb07-May07) – other WP 620-0x000
  - prototype code
  - generate test harnesses
  - iterate

WP 620-01000

Title: Management, configuration management and interfaces		
Provider: MSSL	**Manager**: Simon Rosen	
Start: 01/10/2006	**End**: 01/03/2012	**Effort**: 0.1 MY/year
Objective: To manage the spectral extraction workpackages to ensure that the software development is produced to specification and on time.		
Tasks: This package covers the management responsibilities for finance, reporting, internal organisation, interfacing with the wider GAIA consortium and software development management 1. Communicate with ESA, Meudon (CU6 lead institute) and where necessary, other GAIA CUs, for all CU6-specific and general GAIA software and other project issues (e.g. calibration, archiving, interfacing, scheduling, planning etc.), to ensure timely delivery and quality control of required software. 2. Act as point of contact (POC) for information flow, SPR/SCRs, and CU6/project decisions. Respond to problems/issues. 3. Oversee internal manpower assignments, resourcing, scheduling and budget control. 4. Oversee all design, prototyping, coding, testing and validation/quality control of spectroscopic extraction workpackages for which MSSL is responsible. 5. Ensure documentation is written, maintained and appropriately organised. 6. Ensure local staff are fully familiar with project documentation and procedures. 7. Interface with PPARC (via the Vega consortium) with regard to financial reporting and related issues. Maintain close liaison with Cambridge/Vega consortium relating to UK involvement with the GAIA project.		

WP 620-02000

Title: Detailed functional analysis		
Provider: MSSL	**Manager**: Simon Rosen	
Start: 01/10/2006	**End**: 01/10/2007	**Effort**: 0.3 MY/year
Objective: To produce a detailed functional analysis of the spectral extraction tasks.		
Tasks: This workpackage defines the spectral extraction task from raw data to calibrated spectra. 1. consider and design the workflow for extracting spectra from data within the context of the Astrium design, removing background, cleaning and calibrating them 2. identify and scope the data storage requirements and transfer rates 3. identify external and auxiliary data required 4. define the timescales for reiteration		

WP 620-03000

Title: Extract Spectra		
Provider: MSSL	**Manager**: Simon Rosen	
Start: 01/10/2007	**End**: 01/03/2012	**Effort**: 0.2 MY/year
Objective: To develop the code to extract the spectra from the RVS data stream, together with associated parameters.		
Tasks: This workpackage defines the optimal spectral extraction algorithm: 1. consider the implications of the Astrium-3 design on the extraction concept 2. explore and develop extraction concepts 3. prototype different extraction schemes 4. develop verification code 5. test prototype schemes on simulated data in a standard set of scenarios (LR/HR/mixed windows, overlapping windows) 6. select optimal schemes 7. develop code for optimal scheme 8. verify spectral extraction		

WP 620-03000

Input: 1. GAIA-C6-TN-OPM-DK-001-2: *Gaia spectroscopic processing – Preliminary functional and data flow analysis* 2. Gaia-3 design and operational parameters 3. simulated datasets
Output: 1. Updated functional analysis for spectral extraction: GWP-S-620-02000 simulated data 2. verified code module 3. test datasets

WP 620-04000

Title: Apply calibration		
Provider: MSSL	**Manager**: Simon Rosen	
Start: 01/10/2007	**End**: 01/03/2012	**Effort**: 0.2 MY/year

Objective:
To develop the code to convert the RVS raw spectra from spacecraft-specific to standard units.

Tasks:
This workpackage will produce the code which applies the calibration (determined elsewhere) to the raw data. All calibration aspects need to be considered, in particular the wavelength and the photometric throughput.

1. consider the implications of the Gaia-3 design on the calibration concept
2. provide interfaces for auxilliary data (Gaia and non-Gaia)
3. develop code for calibration at CCD level (bias removal, flat fields, cosmetic defects)
4. develop code for wavelength calibration including propagation of systematic and random errors
5. develop code for photometric throughput calibration including propagation of systematic and random errors
6. develop verification code
7. verify calibration using simulated data

WP 620-05000

Gaia DPAC WP:		GWP-S-620-05000
Title: Model background generated by extended sources		
Provider: MSSL	**Manager**: Simon Rosen	
Start: 01/10/2006	**End**: 01/03/2012	**Effort**: 0.2 MY/year

Objective:
To develop the code to model the background contribution within spectra from extended sources.

Tasks:
This workpackage will produce the code which identifies and accumulates the background flux contribution from extended sources. The main extended source component will be cosmic in origin (galaxies, nebulae) but there may also be internal sources (optical ghosts). The catalog of extended objects to be used will be available from public sources (eg NGC2000, MCG, CGCG).

1. identify limiting values (G_{RVS}magnitudes arcsec^{-2}, extent, emission line strengths, central condensation *etc*.) for consideration as extended sources
2. classify types of cosmic extended sources (e.g. are globular clusters extended sources?)
3. collate appropriate catalogs or image data for cosmic extended objects
4. identify and collate quantitative information on other sources of diffuse background such as Zodiacal light and diffuse galactic emission
5. identify sources of internal extended objects (ghosts, scattered light)
6. develop code to generate internal extended source structure from bright star catalogs
7. develop code to generate spectral contributions from individual extended sources
8. develope code to model other diffuse sources of bacground.
9. develop code to accumulate spectral contributions from individual extended sources and other diffuse emission for observed object windows including propagation of systematic and random errors
10. develop verification code
11. verify extended background prediction using simulated data

WP 620-06000

Gaia DPAC WP:		GWP-S-620-06000
Title: Model background generated by point-like sources		
Provider: MSSL	**Manager**: Simon Rosen	
Start: 01/10/2006	**End**: 01/03/2012	**Effort**: 0.2 MY/year

Objective:
To develop the code to model the background contribution within spectra from point-like sources.

Tasks:
This workpackage will produce the code which identifies and accumulates the background flux contribution from point-like sources. Point sources will be observed with *Gaia* to G $\sim$ 20, but there will not be a complete sample of objects in crowded regions, or for bright objects, so other external information and catalogs will be required.

1. identify the characteristics of the point-source background that needs to be considered (limiting G_{RVS} magnitude *etc*.)
2. identify the extent to which background can be modelled giving limited photometric/spectroscopic knowledge about the objects, and develop a strategy to handle this
3. collate appropriate catalogs for point sources (eg *Hipparcos*, USNO-B1.0)
4. identify access to *Gaia* skymapper, photometry or spectrometry data from earlier scans for use in constructing point source background, together with accurate positions
5. generate a strategy as to what information should be used from which resources under particular conditions (limiting magnitude, extent of *Gaia* data products *etc*.)
6. develop code to generate spectral contributions from individual point sources, taking into account their classification/spectral types
7. develop code to accumulate spectral contributions from individual point sources for observed object windows including (for bright sources) at wavelengths outside the RVS band and also including the propagation of systematic and random errors
8. develop verification code
9. verify point-source background prediction using simulated data

WP 620-07000

Title: Clean spectra		
Provider: MSSL	**Manager**: Simon Rosen	
Start: 01/10/2006	**End**: 01/03/2012	**Effort**: 0.2 MY/year
Objective: To develop the code to subtract backgrounds from the spectral data		
Tasks: This workpackage will produce the code which subtracts the backgrounds from the spectra and removes the cosmic ray contamination. 1. identify optimal algorithms for dealing with cosmic ray identification 2. identify optimal algorithms for dealing with cosmic ray subtraction 3. develop code to perform cosmic removal including propagation of systematic and random errors 4. develop code to subtract extended and point source background including propagation of systematic and random errors 5. develop verification code 6. verify cleaning algorithm using simulated data		

WP 620-08000

Title: Normalisation to the continuum		
Provider: MSSL	**Manager**: Simon Rosen	
Start: 01/10/2006	**End**: 01/03/2012	**Effort**: 0.2 MY/year
Objective: To develop the code to normalise the spectra by their continuum		
Tasks: This workpackage will produce the code which normalises the spectra by their local continuum. This is in preparation for the radial velocity determinations. 1. classify different natures of the spectral continua as a function of object type 2. identify optimal algorithms for robust automated fitting to the continuum, dealing with emission/absorption lines of different number, depth and width 3. develop code to fit the continuum, and to divide by it to normalise the spectra, including propagation of systematic and random errors 4. develop verification code 5. verify normalisation code using simulated data		

WP 620-09000

Title: Validation: spectra extraction		
Provider: MSSL	**Manager**: Simon Rosen	
Start: 01/10/2007	**End**: 01/10/2011	**Effort**: 0.3 MY/year

Objective:
To validate the spectral extraction tasks.

Tasks:
This workpackage validates the entire spectral extraction chain end-to-end (individual components of the chain are verified within their own workpackages).

1. specify the simulated data for testing the spectral extraction chain in different scenarios (nominal windows, overlapping windows, different extended and point-source backgrounds *etc.*)
2. specify and write the verification analysis software
3. verify normalisation code using simulated data
4. analyse end-to-end extraction to quantify and prove capability in different scenarios