The Hubble Spectroscopic Legacy Archive

Molly S. Peeples[1], Jason Tumlinson[1], Andrew Fox[1], Alessandra Aloisi[1], Scott Fleming[1], Robert Jedrzejewski[1], Cristina Oliveira[1], Thomas Ayres[2], Charles Danforth[2], Brian Keeney[2], Edward Jenkins[3]

[1]Space Telescope Science Institute, Baltimore, MD
[2]University of Colorado, Boulder, CO
[3]Princeton University, Princeton, NJ

28 April, 2017

ABSTRACT

With no future space ultraviolet instruments currently planned, the data from the UV spectrographs aboard the Hubble Space Telescope have a legacy value beyond their initial science goals. The goal of the Hubble Spectroscopic Legacy Archive (HSLA) is to provide to the community new science-grade combined spectra for all publicly available data obtained by the Cosmic Origins Spectrograph (COS) and the Space Telescope Imaging Spectrograph (STIS). These data are packaged into "smart archives" according to target type and scientific themes to facilitate the construction of archival samples for common science uses. A new "quick look" capability makes the data easy for users to quickly access, assess the quality of, and download for archival science. The first generation of these products for the far-ultraviolet (FUV) modes of COS was made available online via the Mikulski Archive for Space Telescopes (MAST) in early 2016 and updated in early 2017; future releases will include COS/NUV and STIS/UV data.

Contents

1. Introduction

Data archiving is a crucial component of the operation of an astronomical observatory. Archives ensure the legacy of the observatory, acting as multipliers for its science output by enabling science investigations unrelated to those in the proposals that obtained the data. To maximize their use by the community, archives need to be well populated, easy to use, and designed so that astronomers can quickly determine which targets of a given type have been observed. We describe here a new archival resource for spectroscopic data from the *Hubble Space Telescope*, named the *Hubble* Spectroscopic Legacy Archive (HSLA), hosted at the Mikulski Archive for Space Telescopes (MAST) at STScI (https://archive.stsci.edu/hst/spectral_legacy/).

The HSLA is an effort designed to maximize the longevity of the UV spectroscopic data and to accelerate the scientific study of these data. The HSLA is led from the Institute with involvement from several spectroscopists in the community. The HSLA working group was formed following a workshop held at the Institute in November 2012, *Enhancing the Legacy of HST Spectroscopy*, chaired by Alessandra Aloisi and Stefano Casertano, in which community input was given on the needs for archival products for spectroscopy, and the Institute began planning the tools and products necessary to meet these needs.

Though MAST has long provided combined images to the community (e.g., Fruchter & Hook 2002; Delude & Dressel 2009), the extant Institute pipelines only combine, e.g., COS (Cosmic Origins Spectrograph) spectra up to the visit level, combining spectra taken at different spectral dither positions (FP-POS) but not different cenwaves or gratings (e.g., the G130M and G160M modes). As observing programs are often set up in order to cover a long continuous wavelength range, or different programs might observe the same target in different modes with overlapping wavelength ranges, science-grade co-added spectra would be a top priority enhanced product to add to the archives. Owing to its popularity and more homogenous modes than STIS, we chose to deliver the first set of these products for COS/FUV.

The HSLA consists of two complementary pieces. The first, discussed in Section 2, is the creation of uniformly reduced and co-added spectra (beginning with both the M- and L-mode COS/FUV gratings), combining data taken from multiple programs and epochs to provide higher signal-to-noise and a wider wavelength coverage than the standard **calcos** pipeline. The second, detailed in Section 3, is packaging these data into easily accessible "smart archives" according to target type and scientific themes (such as "solar system," "early type stars," "white dwarfs," and "starburst galaxies") to facilitate the construction of archival samples for new *Hubble* proposals and for

general science usage. We conclude in Section 4 with possible future directions for the HSLA.

Step	coadd_x1d (Colorado-COS IDT)	coscombine (Wisconsin)	counts_coadd (UMass/STScI)	Adopted
Decide which files to combine	input file	input file	named in input file	Automated search of target directory
Wavelength Shifts	cross-correlate strong ISM lines	cross-correlate on manually selected lines	manually derived from single line	**calcos** pipeline wavelengths
Coadd method	nearest-neighbor	linear interpolation	shift & add (equivalent to nearest neighbor)	shift & add
Renormalize	scale to reference	none	none	none
Units of coadd	flux	counts	counts	counts
Data Quality (DQ) handling	discard or deweight	deweight	delete those counts	delete those counts
Background	done by calcos	done by calcos	subtract background from gross counts	subtract background from gross counts
Errors	fix with empirical relation from Keeney et al. (2012)	compute from gross counts, correct for continuum rms	Gehrels (1986)	Gehrels (1986)
Fluxing	yes, with scaling to reference	no scaling	No	apply mean flux curve
Binning	optional	optional	1, 3, or 6 pixels always generated	leave to user

Table 1: Summary of steps taken by co-addition codes already in use in the community (coadd_x1d, coscombine, and counts_coadd), with the method adopted by the HSLA shown in the rightmost column.

2. Science-Grade Co-added Spectra
Previous versions of the COS pipeline only add up to the visit level, e.g., **calcos** combines data taken with the same grating, central wavelength, and aperture but with different FP-POS (i.e., exposure setting), akin to combining over dithers for images taken with a given filter. Many science cases, however, benefit from having a co-added set of *all* data with similar resolution on a given target (exceptions include time-variable sources). Such products have not generally been provided as part of the Hubble archives, though MAST does host community contributions of High Level

Science Products (see https://archive.stsci.edu/hlsp/index.html), which potentially could include such spectral co-additions.

We considered three community codes for co-adding COS/FUV spectra: (1) `coadd_x1d` developed by the Colorado / COS IDT team (Danforth et al. 2010 and Keeney et al. 2012); (2) `coscombine`, developed at the University of Wisconsin (Wakker et al. 2015); and (3) `counts_coadd`, developed by the University of Massachusetts / STScI COS-Halos group (Tumlinson et al. 2013). Surveying the algorithms used by each of these codes, we determined the in-common steps for combining spectra; as shown in Table 1, our adopted method sometimes uses the most common approach, but in some cases we chose to go in a different direction. In particular, the three community codes were primarily developed in order to search for absorption features in the spectra of quasars or UV bright stars; in this regime, there are generally strong intervening interstellar absorption lines that can be cross-correlated in order to shift all of the spectra on the same wavelength grid. The full COS archive, however, contains both continuum-free targets as well as those too nearby to have a sufficient number of ISM absorbers in their spectra, limiting the ability of an automated universal code to easily do this iterative cross-correlation. For this reason, the HSLA simply adopts the wavelengths derived from **calcos**, with no attempt to shift different exposures relative to one another; this assumption may be revisited for future releases.

We chose to do the co-addition in counts rather than flux, following Gehrels (1986), in order to make propogating the photometric errors much cleaner; we note that the errors should not be correlated. We then use the average flux calibration based on the average FLUXFACTOR for the counts going into each bin to convert to flux density. As shown in Table 2, the counts and average fluxfactor are still retained in the final product. We do not combine data from different resolutions (the M and L-modes); thus, some targets have both low-resolution and medium-resolution co-adds available.

3. "Smart" Archives

The HSLA data are packaged into "smart archives" according to target type and scientific themes (such as "solar system," "early type stars," "white dwarfs," and "starburst galaxies") to facilitate the construction of archival samples for new *Hubble* proposals and for general research.

The HSLA provides quick access to the raw and reduced data in tables sorted by target class. Each sample is organized into sortable, searchable tables with basic demographics (such as typical S/N or sample-specific information, e.g., quasar redshift) and thumbnail plots of the co-added spectra. Each target then has its own "quick look" page, which includes plots of the co-added spectra to allow users to assess data quality (see Figure 1). This page also includes a demographic section summarizing the distribution of instrument modes and programs for that target and a light curve showing the evolution of source flux with time, in selected 10Å wavelength bands, allowing for a quick assessment of time-variability at different wavelengths.
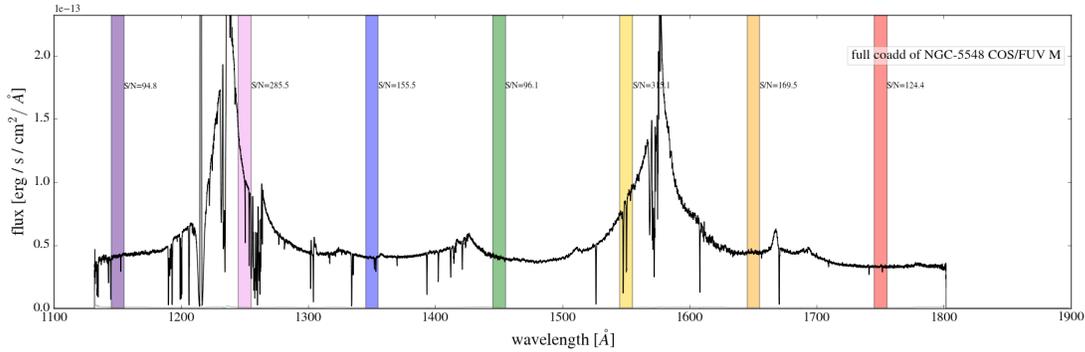
Figure 1: A sample co-added spectrum, smoothed over 6 pixels. The colored bars denote the 10Å windows in which the S/N per resel (taken to be 6 pixels) are calculated for comparison to the individual exposures.

3.1 A guide to the database's contents

Samples: A central facet of the HSLA are the "samples" of targets organized by target type and/or scientific purpose. These samples have been constructed in such a way that they are downloadable and customizable as a unit. To access the samples, visit "Data Access" on the main HSLA webpage (https://archive.stsci.edu/hst/spectral_legacy/) and scroll through the tables. Each individual entry in these tables is a "sample", eg. "Extrasolar Planets" and "Early Type Stars". Everything associated with a sample—including the list of targets and all their coadds—can be downloaded with a single click. Inside the resulting tar.gz file are various tables containing target and exposure information, labeled with the sample name. Of particular interest are the table of targets (STAR_sample) and exposures ("STAR_exposures") and their metadata in .html and .fits formats. (The FITS tables can be found by changing the .html in the url to .fits and/or grabbed with wget.)

In addition to the classification-oriented samples, we have provided a sample including all of the 1624 targets with public data as of March 2017 ("all_targets"), and two subsets: one including all targets with FUV data and another for NUV data. As many targets have both NUV and FUV data, they appear in both samples. The NUV coadds will be available in a future release. Note that the assignment of targets to samples was done according to human judgment informed by visual inspection and literature searches, as the "target descriptions" do not have a direct correspondence to our sample categories.

Target directories: These directories contain all the target-specific files. These include the original **calcos**-produced x1d files that go into the coadds (unmodified since their download from MAST). There is also a catalog of exposures for the target (all_exposures) and the quick look web page and the images linked there. Because the data for individual targets observed under different names have been consolidated, the all_exposures table and the included x1d files for some targets will have multiple entries with different names.

Coadd filenames: The HSLA database includes new high-level coadds of COS FUV data (G130M, G160M, G140L). These coadds are stored in files called "_coadd_[grating]_final_[lifetime].fits. [grating] can be "G130M", "G160M", "G140L", or "FUVM", which is the splice together of the G130M and G160M coadds,

if both are available. Then [lifetime] takes the values "all", which is the sum of all data collected at any lifetime position. Files named with values "lp1", "lp2", and "lp3" combine data collected at that lifetime position only. For targets observed at only one LP (e.g., 2), the data in "all" and "lp2" will be identical.

Coadd contents: We summarize the coadd fits file contents in Table 2. The coadd files include arrays for WAVE, FLUX, and ERROR which will serve most purposes. These default outputs were derived from the counts-rate based coadds. Users interested in exposure-time weighted fluxes should use "FLUXWGT" and "FLUXWGT_ERR". Users interested in the cleanest statistics should adopt "NETCOUNTS" and "NETCOUNTSERR".

Vector	Units	Comment
WAVE	Å	Correspond to "raw" pixels with no binning
FLUX	$erg\ s^{-1}\ cm^{-2}\ Å^{-1}$	Main flux vector, determined from NETCOUNTS
ERROR	"	Main error vector, determined using Gehrels (1986)
FLUXERR_UP	"	Two sided errors on flux
FLUXERR_DOWN	"	Two sided errors on flux
GROSSCOUNTS	$counts\ s^{-1}$	coadded gross count rate
NETCOUNTS	$counts\ s^{-1}$	coadded net count rate
NETCOUNTSERR_UP	$counts\ s^{-1}$	coadded net count rate error (Gehrels 1986)
NETCOUNTSERR_DOWN	$counts\ s^{-1}$	coadded net count rate error (Gehrels 1986)
BACK	$counts\ s^{-1}$	coadded 1D background count rate
EXP_PIX	s	Summed exposure time per pixel
DQ	bitwise sum	Propagated from all exposures, meaning retained
FLUXFACTOR	$erg\ cm^{-2}\ Å^{-1}\ count^{-1}$	exposure-time weighted FLUX / NET from x1ds
FLUXWGT	$erg\ s^{-1}\ cm^{-2}\ Å^{-1}$	exposure-time weighted mean flux
FLUXWGT_ERR	$erg\ s^{-1}\ cm^{-2}\ Å^{-1}$	error in exposure-time weighted mean flux
CHI2	-	χ^2 of all exposures relative to FLUX

Table 2: The detailed contents of the coadd files

3.2 A note about naming conventions (or the lack thereof)

The names of targets in the Hubble archive were those assigned by the developer of the Phase 2 observation specification form, often leading to multiple names for the same target (e.g., PG-1630+337, PG1630+337, and [HB89] 1630+337 all refer to the same target). When a decision had to be made, we adopted the most conventional name in our judgment, usually following SIMBAD for Galactic objects and NED for extragalactic targets. In cases where a target has only been observed once, we defaulted to name assigned by the PI (e.g., "LY00"). As the HSLA becomes integrated into MAST, we expect for naming conventions to become both more automated and more consistent.

3.3 A note about the lightcurves

The quicklook pages include at the bottom lightcurves (plots of flux versus time) in each of the 10Å bands used to calculate the S/N per resel in the individually plotted spectra. For most targets, the time histories are fairly uninteresting: the source was not variable over the relevant timescales. For targets with more complex temporal behavior, users may want to think twice about whether the co-additions will be able to adequately serve their research objective.

3.4 Spectator: a tool for generating COS quicklooks

The HSLA tables, html pages, and plots are generated using the custom-written, publicly-available python code, Spectator (http://github.com/peeples/Spectator). This code can be easily modfied for quickly viewing custom samples of COS data.

4. The HSLA Going Forward

The initial release of the HSLA was in February 2016, in time for use in preparation of Cycle 24 proposals, with a preliminary version showcased at the January 2016 AAS meeting in Kissimmee, Florida (Peeples et al. 2016). In March 2017, in advance of the Cycle 25 *Hubble* proposal deadline, we updated the HSLA to include more than 200 new targets and recent public data on targets included in 2016's release. Importantly, all data have been re-calibrated to include the new COS FUV dispersion solution. Future releases of the HSLA will contain co-added COS/NUV data and STIS UV data, with any updates to the co-addition algorithm or references files associated with data included in previous releases. Currently the HSLA is a static archive, but our aim is for the HSLA to be fully integrated into MAST, including co-added spectra available with fresh data and automatic updates for newly available data and/or any changes to the **calcos** or **calstis** calibration pipelines, such as improvements to the wavelength-scale dispersion solution and changes to the reference files. These evolutionary enhancements will ensure the scientific legacy of *Hubble*'s spectroscopic data well past the lifetime of *Hubble* itself.

Acknowledgements

Change History for COS ISR 2017-04

Version 1: 28 April 2017 – Original Document

References

Astropy Collaboration, 2013, A&A, 558, 33

Danforth, C., et al., 2010, ApJ, 720, 976

Dulude, M. and Dressel, L. 2009, Instrument Science Report WFC3 2009-11, "First-Order Test of WFC3/IR Imaging: from Phase 2 Proposal to MultiDrizzle Product"

Fruchter, A. and Hook, R. 2002, PASP, 114, 144

Gehrels, N., 1986, ApJ, 303, 336

Keeney, B., et al. 2012, PASP, 124, 830

Peeples, M. S., et al., 2016, AAS, 227, 444.01

Tumlinson, J. et al., 2013, ApJ, 777, 59

Wakker, B. P. et al., 2015, ApJ, 814, 40