

Method for boosting dispersive spectrograph stability 1000× using interferometry with crossfaded pairs of delays

David J. Erskine*

Lawrence Livermore National Laboratory, Livermore, California, United States

Abstract. We demonstrate a key step along a technical route to achieving cm/s scale accuracy for astronomical spectrographs over long (multi-year) time scales, which is critical for the Doppler characterization of Earth sized exoplanets, and measurement of small cosmic redshift drift over many years. This same technique also enables searching exoplanet atmospheres for biosignificant molecules in direct planet imaging using, otherwise, insufficiently low resolution and drift prone dispersive (grating or prism) spectrographs. Using a method called crossfading for externally dispersed interferometers (EDIs) to get highly robust spectra, we recently demonstrated a factor of 1000× reduction in the net shift of an EDI measured ThAr line to a deliberate simulated wavelength translation of the detector. This 1000× gain in disperser stability can be combined with conventional stability gains afforded by fiber scramblers, vacuum tanks, and thermal control, to provide an additional 1 to 3 orders of magnitude reduction in the net point spread function shift drift. Crossfading combines high- and low-delay fringing signals that react oppositely in phase to cancel their net reaction to a detector wavelength drift. This can be implemented by an interferometer addition to a facility spectrograph. © *The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: [10.1117/1.JATIS.7.2.025006](https://doi.org/10.1117/1.JATIS.7.2.025006)]

Keywords: high-resolution spectroscopy; doppler radial velocimetry; exoplanet detection; cosmic redshift drift; laser frequency comb calibrator; externally dispersed interferometer; dispersed fixed delay interferometry; spectrograph stabilization.

Paper 20166 received Nov. 11, 2020; accepted for publication May 20, 2021; published online Jun. 10, 2021.

1 Synopsis: Crossfading-EDI Robust to Spectrograph Drift

A significant instrumental error for dispersive (grating or prism) spectrographs is uncontrolled drift Δx of its focal point [point spread function (PSF) or lineshape function (LSF)] relative to its detector pixels. This can prevent achieving photon or detector limited performance¹ for high-resolution spectroscopy and Doppler velocimetry in a variety of science and engineering fields. A hybrid technique of externally dispersed interferometer (EDI)^{2–14} has been used as an inexpensive and compact alternative for Doppler and high-resolution spectroscopy and has been used by other researchers to find exoplanets^{15–17} around stars HD102195 and HD87646 in 2005 and 2016. The EDI can be formed from most spectrographs, after their construction, by insertion of a small interferometer in series with its input beam.

The EDI was already known for its excellent stability properties. We have recently discovered improvements in the apparatus design and data processing called “crossfading” that dramatically improve its stability to a drift Δx by to one to three orders of magnitude. Figure 1 shows a demonstration producing a 1000× reduction of reacting shift to a Δx on a measured ThAr spectral lamp line whose data are artificially drifted—the EDI output (red peak) is virtually stationary. The crossfading technique uses pairs of interferometer delays, rather than a single delay (as used by the basic EDI) as well as strategically chosen weights during data processing that cancel error signals among each delay pair. We envisage improving the resolution and stability performance of facility spectrographs by introducing small interferometers prior to their

*Address all correspondence to David J. Erskine, erskine1@llnl.gov

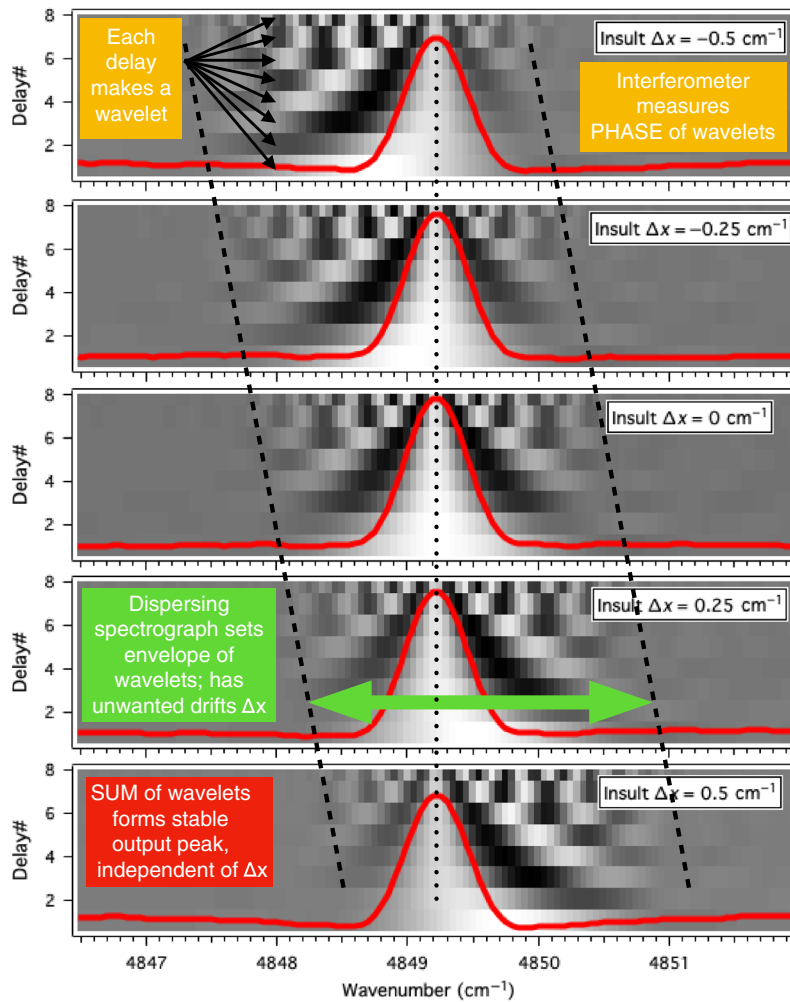


Fig. 1 Synopsis: using an EDI with multiple overlapping delays, one can strategically choose weights in the wavelet sum (grayscale patterns) that forms the output spectrum (red peak) to cancel the effect of unwanted drift (insult Δx) of the dispersing spectrograph. This simulation uses an artificially imposed drift on measured EDI data of a ThAr lamp at the Mt. Palomar Hale Telescope.¹¹ Each delay creates a wavelet when its moiré signal is restored during processing to its original high frequency. The wavelet envelope shape and position are dictated by the dispersive spectrograph instrument response; note it shifts as Δx changes from -0.5 to 0.5 cm^{-1} (from top to bottom panels). Remarkably, a reduction by $\sim 1000\times$ in shift of final ThAr peak position is achieved. The technique, called “crossfading EDI,” can be combined with conventional mitigations (fiber scramblers, thermal control, and vacuum tanks) to multiplicatively boost the net spectrograph stability from one to three orders of magnitude beyond using conventional mitigations alone.

entrance slit. The increased data processing complexity and the increased flux losses due to added optical surfaces are penalties but outweighed by the dramatic improvement in signal-to-noise ratio (SNR) due to decrease in the instrumental noise. This can improve both Doppler velocimetry and high-resolution spectroscopy applications.

The crossfading EDI advantages can be combined with conventional mitigations (fiber scramblers, thermal control, vacuum tanks, etc.) to multiplicatively boost the net spectrograph stability several orders of magnitude beyond what can be achieved by conventional mitigations used alone.

2 Motivation: Spectrograph Drifts

Modern astrophysics demands extremely stable wavelength measurements, such as few cm/s scale Doppler radial velocimetry for Earth-like planet detection¹ and multi-year cosmic redshift

drift measurements.^{18,19} Stable measurements are needed for remote sensing for identifying molecular species spectroscopically on an airborne or space-borne platform enduring the vibrations and accelerations of flight while still being compact and lightweight. Raman spectrometers identifying molecules for biomedicine can suffer from drift when engineered for compactness and portability rather than stability.

Dispersive spectrographs (grating or prism) are often used to make these measurements. The chief instrumental challenge is often not spectral resolution in resolving a feature, or large photon noise due to insufficient flux, but lack of wavelength stability. Fischer et al.¹ described some astrophysical high-resolution spectrographs whose low noise performance has a floor limited by instrumental errors rather than photon noise (e.g., their Fig. 5). The PSF of conventional spectrographs can drift in relative position (Δx) and shape under thermal changes to the diffraction grating or detector, fluctuations of air internal and external to spectrograph, a changing pupil or gravity vector as target moves across the sky, and flexure of optical fibers (creating mode changes in profile) if fibers are used to communicate the light.

An example of a difficult PSF drift error is the time-dependent stack of spectra shown in Fig. 35 of Ref. 11, from the near infrared (NIR) TripleSpec spectrograph of our T-EDI project (used in demonstrations described below). This shows a Δx that varies rapidly spatially and irregularly with both polarities along the dispersion direction. This makes it impractical to remove by simple monolithic translation of the spectra, which is a conventional means of correction during analysis and postdetection.

Conventional means for mitigating drift from occurring in the first place include vacuum tanks, thermal insulation and control, adaptive optics, and fiber-optic scramblers. Often the former are bulky or heavy and precluding airborne or space-borne platforms. These conventional means reduce the “insult” Δx , which is the drift appearing at the detector. Let the error in the processed output spectrum be $\delta\lambda_{\text{out}}$ or, in our preferred dispersion variable for interferometry, wavenumbers $\delta\nu_{\text{out}}$ rather than wavelength λ , where $\nu = 1/\lambda$ in cm^{-1} and $\delta\nu = -\delta\lambda/\lambda^2$. For conventional dispersive spectrographs, $\delta\nu_{\text{out}} = \Delta x$ since the output spectrum is directly linked to the position of features on the detector. (This is different from the interferometer, as we shall see.)

The use of spectral calibrants (iodine cells and ThAr lamps) simultaneous with the input stellar spectrum is very important but by itself may not completely remove an error, because the spectrograph optical behavior and/or detector pixel position may not be perfectly uniform along the dispersion axis.

A laser frequency comb spectral calibrant is useful but typically expensive. Importantly, because the laser calibrant adds to the input spectrum rather than multiplies it (such as an iodine vapor absorption cell), it is difficult to ensure that the calibrant illuminates exactly the same pixels detecting the stellar spectrum. Mismatch in illuminating different pixels detecting calibrant and stellar spectra can lead to significant errors. For example, the 3-cm/s needed to measure the 10-cm/s motion of an Earth-like exoplanet corresponds to 10^{-5} pixel, given a typical 3 pixels per 10 km/s stellar linewidth scaling of spectrographs. Second, there are analysis difficulties when an additive calibrant is steady in amplitude while the input spectrum fluctuates, especially since the fluctuation could be ν dependent. Finally, the additive nature of the laser calibrant does not produce the heterodyning effect (producing moiré patterns) that boosts the effective spectral resolution of the spectrograph, as does the interferometer we advocate.

A particularly pernicious error type is irregular placement of detector pixels (pixel registration error) from manufacturing variability [Fig. 2(a)]. This forms an insult Δx , which can vary with wavenumber and be irregular and bipolar. These misplacements can be milli-pixel in magnitude (corresponding to 300 cm/s) and thus be limiting for Doppler velocimetry of small planets. It is extremely challenging technically for dispersive spectrographs to achieve 10^{-5} pixel stability over years time scale.

This irregular spatial pixel misplacement is not mitigated by fiber scrambling, thermal control, or vacuum tanks (although these mitigate other types of insults). This error can potentially also vary with time if during a different epoch the entire detector is accidentally displaced, or when the entire spectrum is shifted by a different barycentric Doppler component, such as for observations 6 months apart. Cosmic redshift drift measurements and Doppler measurements of Earth-like planets at Earth-like orbit sizes require making cm/s scale measurements over several

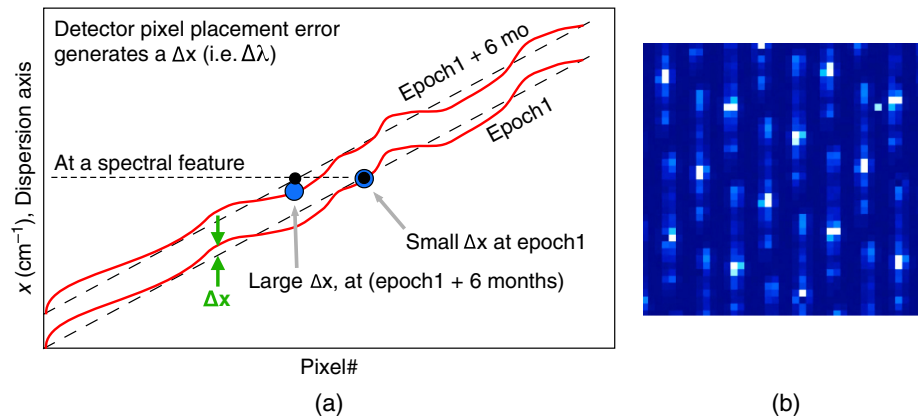


Fig. 2 (a) A pernicious type of error is misplacement of detector pixels during manufacture since it affects spectrographs that are otherwise optically ideal. The insult Δx can vary irregularly and in bipolar manner with wavenumber and is not removed by conventional mitigations such as fiber scrambling. The Δx can vary with date from a large barycentric Doppler component as the Earth revolves around the Sun over 6 months, or if different rows of a detector are used, or if a detector fails and is replaced. Fortunately, this error type can easily be corrected by crossfading since the same “baked in” Δx pattern is automatically applied to each delay of a sequential multiple-delay interferometer such as T-EDI.⁹⁻¹² (b) An example challenging drift problem at the GPI IFS²¹ (small snippet of data shown). This has many close packed spectra, one for each spatial portion of a target. Due to the low resolution and sparse density of pixels, the ~ 1 pixel drift due to changing gravity vector²² would generate a significant wavelength shift. This is a barrier to using GPI to characterize exoplanets at high-resolution spectroscopically but with which EDI can help regarding both spectroscopy¹³ and stability.

year timescales. Over these long timescales, it is possible detectors can fail and need to be replaced, and the replacement detector will have different pixel misplacements having a different error Δx .

An example potential application of a crossfading-EDI is the spectrograph for the Gemini Planet Imager (GPI), which analyzes light from direct imaging of exoplanets and other objects. The spectrum is dispersed over very few (~ 20) pixels [Fig. 2(b)] at very low (~ 50 to 100) resolution.²¹ Drifts in wavelength calibration²² due to the changing direction of the gravity vector (as the telescope follows the object in the sky over long exposures) can be ~ 1 pixel. This is a barrier to using GPI to characterize exoplanets at high-resolution spectroscopically, but EDI can help regarding both the spectral resolution¹³ and stability.

3 Solution: Externally Dispersed Interferometer Using Pairs of Delay

Although we recommend using conventional mitigations when affordable, we propose that dispersive spectrograph stability can be further improved by inclusion of a Michelson interferometer in series, which is EDI [Fig. 3(a)]. In the frequency domain, this has an instrument lineshape Fig. 3(b) having a peak shifted to high frequency from zero by an amount set by the interferometer delay, and in the dispersive domain the lineshape is a wavelet [Fig. 4(a)]. Using multiple delays of different values and summing the results, we can produce a net response in dispersion space that is a narrow peak and measure high-resolution spectra at much higher resolution than the disperser used alone. Figure 5 shows an example of the EDI (red curves) resolving a doublet of a spectral lamp (thin black curve) not resolved by the disperser alone (green dashes).

When working with interferometry, the dispersion variable wavenumbers ν is preferred over wavelength λ since the interferometer transmission $T = (1/2)(1 + \cos 2\pi\tau\nu)$ is periodic in ν , where τ is the interferometer delay or path length difference in cm. (The τ has a slight dependence on ν due to changing refractive index of the glass etalons used to form the delay. This has been measured in Fig. 17 of Ref. 11 and calibrated into our analysis.) Also conveniently, when

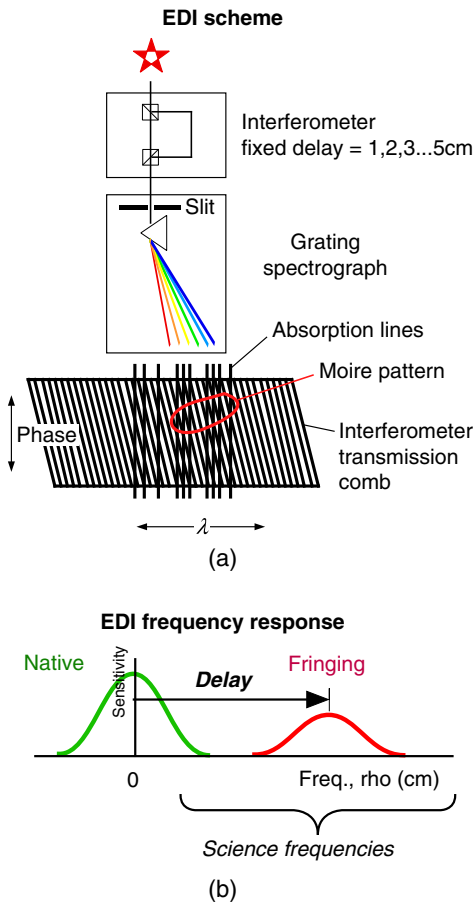


Fig. 3 (a) EDI scheme: a fixed delay interferometer is crossed with a disperser. The sinusoidal transmission of the interferometer creates moiré patterns whose phase encodes Doppler velocity and shape encodes the spectrum's shape (after shifting up in frequency). (b) Heterodyning (multiplication of a signal by a sinusoid) shifts the native spectrograph sensitivity peak (green peak at zero) up to a higher frequency peak (red) where science frequencies typically reside, by an amount equal to the interferometer delay τ , and halved in amplitude. Frequency in units of features per wavenumber (cm^{-1}), conveniently has units of delay, units of cm. Figures reproduced with permission from Ref. 11 by SPIE.

working with wavenumbers (cm^{-1}), the Fourier variable, which is frequency ρ , features per cm^{-1} , has the same units as delay τ in cm. Hence the horizontal axes of some of our graphs are labeled delay, and others are labeled frequency; they are interchangeable.

3.1 Pairs of Delay Cancel Drift

We have recently made improvements in the processing of multiple delay data that produces dramatic increases in the net stability of the output spectrum, robust to an unwanted drift Δx along the dispersion direction of the detector relative to the focal spot of the spectrograph. This improved EDI is called a crossfading EDI, or X-EDI, because at its kernel are strategically chosen weights of a pair of delays that overlap in delay or frequency space. As one moves from one delay to another, the ideal weights change versus ν in an approximate X-shape, hence the name.

The key idea of crossfading (Fig. 6) is that two slightly different delays create moiré patterns of opposite slopes. (In this figure, it is made especially apparent by a feature having a periodic set of lines so that in frequency space it is concentrated in between high and low delays.) An insult drift Δx acting on the moiré slope would create a phase error. The two moirés have opposite slopes—so by combining them it is feasible to cancel their net error effect.

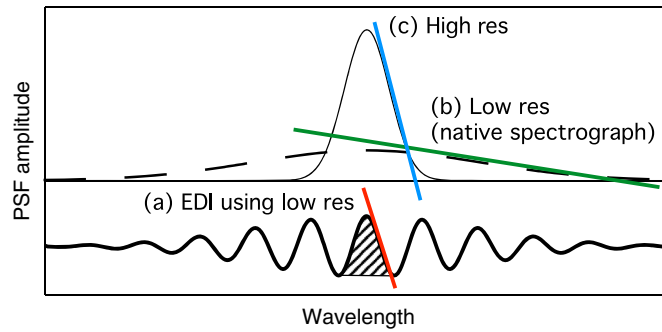


Fig. 4 (a) The EDI has an instrument response in dispersion space (wavenumber or wavelength) in the shape of a wavelet [i.e., FT of frequency response Fig. 3(b)]. The interior shape is very stable, being a sinusoid having only three degrees of freedom (phase, period, and amplitude), which makes it ideally suited to making precision wavelength measurements. In contrast, (b), (c) a conventional spectrograph instrument response has 100s or more degrees of freedom, at least one per grating groove, which makes it more difficult to make an accurate wavelength measurement. For the EDI, the steep slope of the instrument response desired for high Doppler sensitivity is supplied by the slope of an interior sinusoidal fringe [red line in (a)], controlled by the interferometer delay, and this slope can be similar to a high-resolution conventional spectrograph [blue line in (c)]. The envelope of the EDI wavelet shape is the same as the conventional spectrograph [dashed curve in (b)] with which it is used in series. Figure reproduced with permission from Ref. 11 by SPIE.

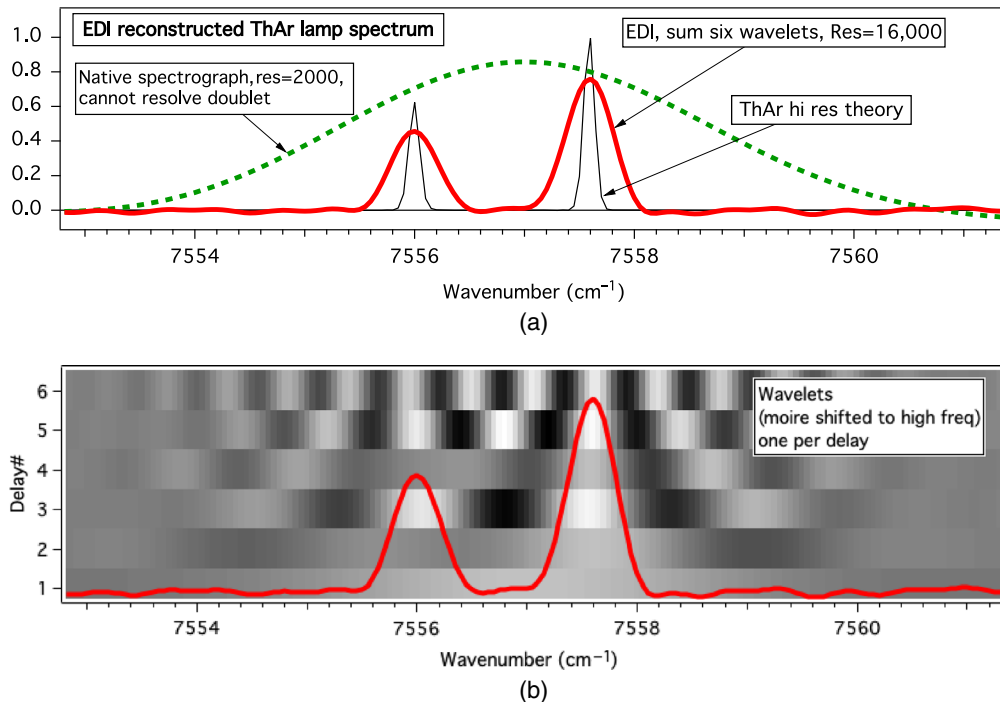


Fig. 5 Demonstration of 8x resolution boosting measuring a ThAr lamp at the Hale Telescope. (b) Phase stepped exposures are made at each of six delays, produce fringing (moiré), and non-fringing (ordinary) spectra. The moiré are shifted to higher frequency in Fourier space by the amount of each delay, to form a set of wavelets. A neighboring ThAr line (off graph at 7532 cm^{-1}) provides interferometer phase alignment as the channels are added to form the output (red curves). (a) The non-fringing spectra at $R = 2000$ (green dash) cannot resolve the doublet (thin black, hi-res theory). The EDI reconstructed spectrum (red curve) equalized to $R = 16,000$ fully resolves the doublet.

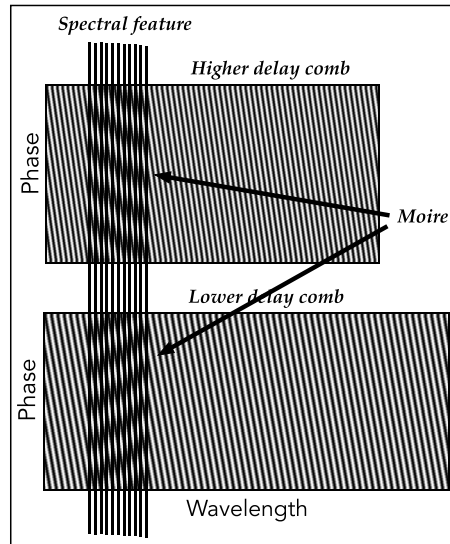


Fig. 6 The crossfading idea has at least two slightly different delays so that the moiré patterns have opposite slopes, cancelling the susceptibility to Δx drift, especially for blended features. This works for the range of frequencies in between each delay of each pair, and when strategically designed weightings are used that account for the instrument lineshape.

Figure 1 of the synopsis shows a simulation of this crossfading process applied to actual data of a multiple-delay EDI measuring a ThAr lamp line (at the 5-m Hale telescope at Mt. Palomar Observatory in the T-EDI project^{9,11,12}). The wavelets' envelope is set by the disperser and we have applied a simulated drift Δx . The fringes inside the wavelets are set by the phase of the measured moiré patterns, and they are mostly independent of the envelope position. (The weights are chosen to zero out the slight dependence to the envelope position Δx .) The output spectrum is the sum over wavelets and is very stable. Its motion is about 1000 times less than Δx , which is remarkable.

This crossfading method can work with a Doppler-measuring EDI (formerly having a single delay) or a spectroscopic-EDI (already having multiple delays). For the Doppler EDI, we modify it to have at least two delays that overlap in delay space. The spectroscopic EDI would continue to use the same multiple delays it already use but use an algorithm we describe below to combine the data, choosing weights in a pairwise manner that cancels the net drift.

3.2 Translational Reaction Coefficient

In the EDI method, the detailed λ or ν determination is decoupled from the spatial scale in the disperser detector and its drift Δx . Instead, ν is determined by the phase of a fringe (intensity measurement) in an interferometer cavity, which is calibrated by a spectral reference such as an iodine cell, ThAr lamp, or laser frequency comb. The cavity PSF is sinusoidal and has only three degrees of freedom (amplitude, period, and phase). This is much easier to control or calibrate than the hundreds of degrees of freedom for a disperser PSF (at least one per grating groove):

Hence, the output drift is given by

$$\delta\nu_{\text{out}} = \Delta x * \text{TRC} = \Delta x / G_{\text{edi}}, \quad (1)$$

where we define translational reaction coefficient (TRC) as the output shift per input shift $\nu_{\text{out}}/\Delta x$ for the spectroscopic method used, and call its reciprocal a stability gain $G_{\text{edi}} \equiv 1/\text{TRC}$. For dispersive spectroscopy, $\text{TRC} = 1$ by definition.

The $\text{TRC} \ll 1$ for EDI because it does not use the feature's position along the spectrograph detector dispersion axis to determine its detailed wavenumber. Instead, it uses the interferometer—specifically it uses the phase of the moiré pattern. The EDI spectrum output is decoupled from the dispersive spectrograph. Significantly, for EDI using multiple delays, we introduced in 2016

(Sec. 10 of Ref. 11) a method called “crossfading,” which theoretically can make $\text{TRC} = 0$ for small insults (relative to the spectrograph resolution element). When Δx is small, the fringe phase shifts are small, and the phasor $e^{i2\pi(\Delta x)(\rho-\tau)}$ that rotates the complex value of the Fourier transform (FT) of the EDI output, at a frequency $\rho \approx \tau$ near a delay, is approximately linear in Δx .

In that previous work, we showed that the benefit of multiple delays, but without special weights, is to achieve in simulation on actual data on a single ThAr line a $\text{TRC} \sim 1/20$ (Fig. 42 of Ref. 11). More recently in 2019²³ and discussed in detail in this report, we further reduce the TRC to $\sim 1/1000$ on the same ThAr line, using strategically chosen weightings to force phase cancellation between each pair of delays for a set of overlapping delays and making other algorithmic improvements. We find this quite exciting. But before we can describe the details of the crossfading technique, we need to provide more explanation of how an EDI behaves and its background since most readers will be unfamiliar with it, and its behavior is not intuitive.

4 Single-Delay EDI Background

The EDI technique has been used for precision Doppler radial velocimetry and high-resolution spectroscopy,²⁻¹⁴ and other researchers using the technique¹⁷ (that they call “DFDI”) have discovered exoplanets^{15,16} around stars HD102195 and HD87646 in 2005 and 2016. Other groups have worked on improving the interferometer design. Wei et al.²⁴ constructed an EDI using a Sagnac style interferometer rather than Michelson design. Because the Sagnac uses counter propagating beams sharing the beam path, it can have a more stable delay value, which has practical benefits. We use a spectral calibrant to compensate for any small [$<\lambda/4$] drifts in delay during analysis since the interferometer phase drift at one ν can be linked to any other ν . We have used both passive and active cavity stabilizations on different apparatuses, to ensure that the delay does not drift more than about $\lambda/4$ to preserve the net fringe visibility over the time exposure, which would diminish the SNR.

The EDI technique is mathematically related to but differing in its practicalities to dispersed Fourier transform spectroscopy (FTS).²⁵ EDI differs in behavior from internally dispersed interferometer techniques such as spatial heterodyne spectroscopy,²⁶ having gratings internal to the interferometer, which can have extremely high resolution over a more limited bandwidth. In contrast, EDI’s bandwidth is not limited by the interferometer and extends over the native spectrograph’s bandwidth.

Figure 3(a) shows the EDI scheme, where an interferometer with a selectable fixed delay is in series with a dispersive spectrograph. The interferometer has a periodic sinusoidal transmission, which multiplies the input stellar spectrum. This embeds a type of calibrant grid on the input spectrum, which is automatically detected by the same pixels as the input spectrum. That is, this interferometer calibrant is married to the input spectrum similar to the absorption spectral reference technique (such as an iodine cell), and when the path of the starlight changes on to different pixels, the interferometer comb naturally follows. (Although, there is still the issue of how to calibrate the interferometer delay sufficiently well. When we used an emission source such as a ThAr lamp on the T-EDI project, we had to alter the design to eventually use collimated beams to improve the overlap between starlight and ThAr. Notably, in the crossfading simulation, we are comparing one ThAr lamp line to another, so this overlap issue is moot for this case.)

Fringing spectra exposures contain simultaneously both the fringing component and the ordinary non-fringing spectrum. By taking data in several exposures while the phase is varied versus time (as for echelle spectrographs) or by varying the phase spatially along the spectrograph slit in a single exposure (as for single-order spectrographs), the data are taken in a so-called phase stepped manner. During analysis, adding the exposures cancels the fringes and yields the ordinary spectrum. Subtracting the exposures cancels the ordinary spectra and reveals the purely fringing spectrum, which are represented by a complex wave. Hence, both ordinary and fringing signals are obtained simultaneously from an EDI.

Multiplication of the sinusoidal transmission comb times the input spectrum creates a down shifted heterodyned component, which are seen as moiré patterns. The frequency shift is set by the delay in cm. (Frequency and delay have the same units here.) High-frequency signals, which

normally would be too fine for the native spectrograph to resolve, are now detectable through their moiré patterns, which are much lower in frequency and thus resolvable. Figure 3(b) illustrates that effectively, the sensitivity peak of the instrument (green peak centered at zero) has been shifted up (red peak) by amount of the delay τ . The FT of this red peak at high frequency is the EDI instrument response in wavenumber or dispersion space, which has a wavelet-like appearance (Fig. 4).

Figure 7 shows an early EDI built by the author in a 1998 project, producing stellar, solar, and iodine moiré data (Fig. 8). For expediency, we used beams in open air without temperature, pressure or convection control, and the native spectrograph and interferometer were subject to thermal-mechanical drifts. Because the iodine absorption spectrum is measured simultaneously with the solar or stellar spectra, the detailed value of the interferometer delay can be determined, and one does not need to stabilize the interferometer cavity better than a rudimentary quarter fringe (to prevent fringes from washing out during long exposures). Under minor drifts of the delay, both the iodine and stellar moiré patterns shift phase by the same amount, and one subtracts them to yield a Doppler measurement that is fairly independent of the precise delay value.

Thus it was practical to obtain m/s scale precision of this early EDI in spite of lack of traditional stabilizing mitigations of vacuum tanks and thermal controls. A 12-m/s amplitude component of the moon tugging the Earth was measured over a month using sunlight from a rooftop fiber-optic heliostat compared to an iodine cell spectral reference, with ~ 8 m/s scatter.⁴ Doppler velocity reproducibility was tested using a back-lit bromine absorption cell as a proxy for a stellar spectrum and was repeatable of few m/s over 11 days.^{3,27} These results are remarkable from a conventionally unstabilized instrument. We expect that new inclusion of (a) cross-fading mitigations and of course (b) convention stabilizing mitigations would improve the precision even further.

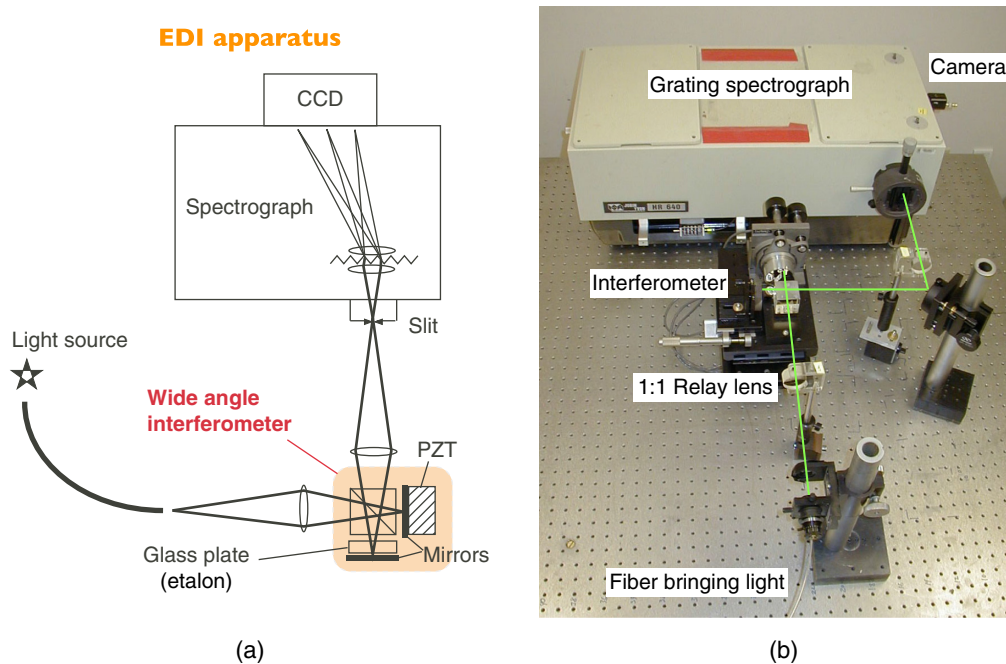


Fig. 7 (a) Schematic and (b) photo of an early single-delay EDI built by a 1998 project at Lawrence Livermore National Laboratory from off-the-shelf components and tested on sunlight,⁴ a stationary backlit bromine absorption cell,³ and starlight.²⁷ Even though it had no environmental controls (open air optical paths), it achieved stability, measured by comparing backlit bromine against iodine absorption spectra, of few m/s over 11 days,^{3,27} and measured the 12-m/s amplitude tugging of the moon on the Earth with 8 m/s scatter over 1 month using sunlight.⁴ The dispersive grating spectrograph was a Jobin-Yvon HR640 (0.7 m length) with maximum $R \sim 20,000$ into a 13-nm band near 513 nm. The entrance slit was often widened to accept more light but at lower resolution. The moiré phase can still be reliably measured at lower disperser resolutions. Figure reproduced with permission from Ref. 6 by SPIE.

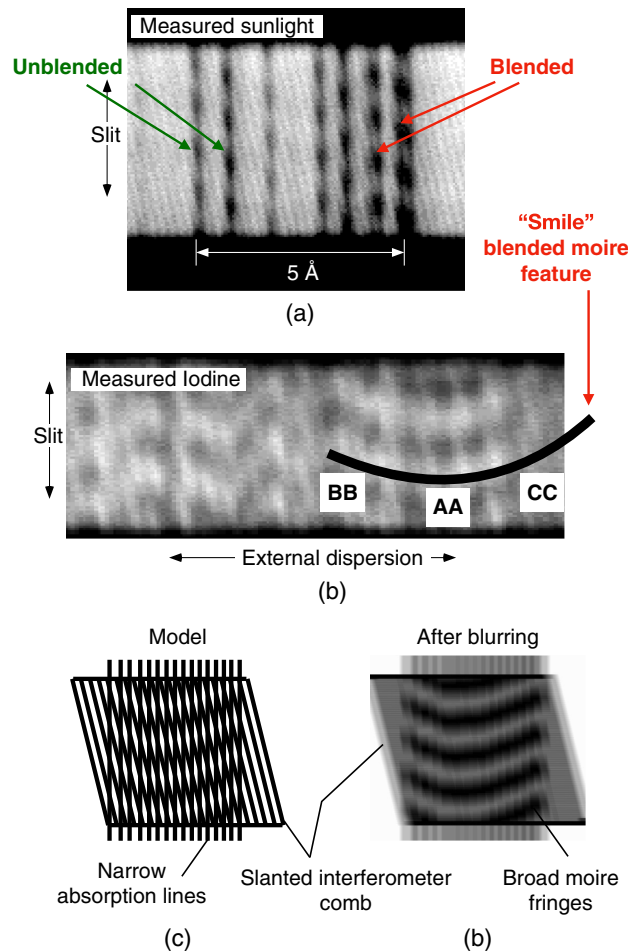


Fig. 8 Fringing spectra measured by the 1998 EDI apparatus (Fig. 7) on (a) sunlight and (b) iodine reference spectra, showing moiré patterns. Solar snippet is near 508 nm, iodine snippet near 540 nm, and wavelength increases to the left. Interferometer phase is along vertical axis. (c) Unblurred and (d) blurred model of iodine line group. During velocimetry observations solar and iodine would be measured simultaneously on the same pixels. Note the smile-like blended feature (BB to CC) in the iodine moiré pattern due to many nearly periodic closely spaced lines, similar in frequency to the 1.1-cm delay interferometer comb (at AA) but slowly changing. Isolated solar lines [left side of (a)] are very robust to spectrograph drift Δx because the interferometer comb and solar spectra move together. However, blended solar lines [right side of (a)] have an apparent moiré slope and thus would create a phase error for a Δx . Similarly for inclined portions of the blended iodine smile-feature in (b). Being bipolar, most of the phase error cancels out when integrated over the band for a Doppler measurement, but there will be a small remainder, which is undesirable. Second, the phase error is not uniform, and thus may couple to any non-uniformity in the spectrograph. In spite of this, the 1998 single-delay EDI performed remarkably well, in the few m/s regime.^{4,27} The crossfading method we proposed can cancel this type of error, by measuring Doppler velocities with two slightly different delays, foretelling even better performance (based on Fig. 5 of Ref. 4) © The Astronomical Society of the Pacific. Reproduced with permission from Ref. 4 by IOP Publishing. All rights reserved.

5 Implementation Examples

5.1 Uni-Phase versus Multi-Phase Recording Styles

Figure 9 shows the difference between multi-phase (a) and uni-phase (b) recording styles. The earliest EDI used multi-phase, which produce slanted fringes, and the phase varies linearly across the spectrograph slit. The slanted fringes are obtained by tilting either the interferometer beamsplitter (BS) or the end mirror of an arm. The multi-phase mode is much easier to

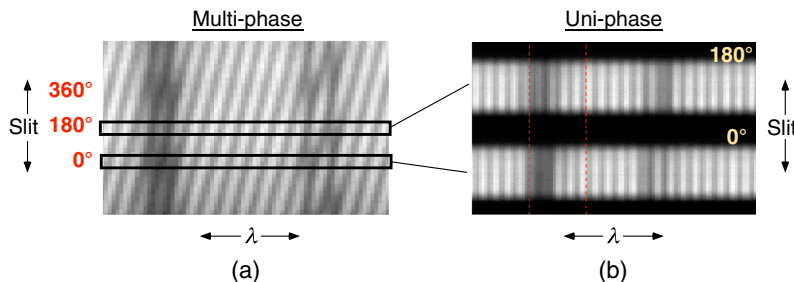


Fig. 9 (a) Multi-phase and (b) uni-phase recording. In multi-phase EDI, all interferometer phases are recorded in a single exposure because the fringe phase is varied (slanted) across the spectrograph slit. In uni-phase recording, the fringes are uniform over the slit and different phases are recorded in different exposures. Uni-phase recording with zero slit-height allows for imaging and echelle spectroscopy.

understand intuitively since the moiré patterns are immediately visible in a single exposure. However, because multi-phase take at least several pixels vertically (along the slit), they cannot easily be used with echelle spectrographs or imaging spectroscopy along the slit dimension. Uni-phase recording [Fig. 9(b)], in contrast, can support a 1-pixel high beam and thus is compatible with echelle or imaging spectroscopy.

5.2 Single-Delay EDI at Lick Hamilton Echelle

In 2002, we tested a single-delay EDI on the Lick Observatory Hamilton echelle spectrograph ($R \sim 50k$) and achieved a resolution boosting of $\sim 2\times$ (to $R \sim 100k$) on starlight using a 3.3-cm delay, without altering the spectrograph or using a narrower slit. The interferometer (Fig. 10) was temporarily inserted into the beam between the telescope and spectrograph entrance slit by mounting on a 30-cm square plate. An even number of reflections on the plate (e.g., five mirrors M1 to M5 and BS) simplified alignment by making it insensitive to slight misorientation.

5.3 Complementary Outputs Available

An ideal interferometer does not absorb light but redirects it into two complementary phased outputs. Figure 11(a) shows that both outputs can be directed toward a spectrograph but necessarily detected on different pixels (otherwise the fringes cancel).

Figure 11(b) shows monolithic prism interferometer designs. Such prisms could have different delays, so instead of swapping out a separate glass etalon to change the delay, one could swap out whole interferometers. This might have the advantage of not upsetting interferometer cavity alignment during delay change since that is fixed in glass dimensions.

6 Some EDI Theory as High-Resolution Spectrograph

6.1 Synthetic Output Spectrum Distinct from Detector Spectrum

Figure 12 compares the instrument lineshapes of three spectroscopy techniques: (a) and (b) purely dispersive, (e) and (f) purely interferometric (FTS), and (c) and (d) EDI, which is a hybrid of the two. The interferometer has a single-pixel detector (e) that accepts all wavenumbers, and from a phase stepped set of at least three intensities generates an ultra-wide sinusoid $\cos 2\pi\tau\nu$ to its output (f) ($\tau = 1.75$ cm). The output is necessarily rebinned to a finer bins to hold the high-frequency information. Inside the linewidth of the dispersive spectrograph, EDI acts like an interferometer, accepting phase stepped intensities (a) to generate a sinusoidal wavelet in the output (b) of the same periodicity as as the pure interferometer but with an envelope $PSF(\nu)$ set by the dispersive spectrograph.

Note if we include the non-fringing component with the fringing, then the EDI instrument response is $[1 + \cos(\cdot)]$ times the dispersive lineshape envelope, instead of $[\cos(\cdot)]$ times the

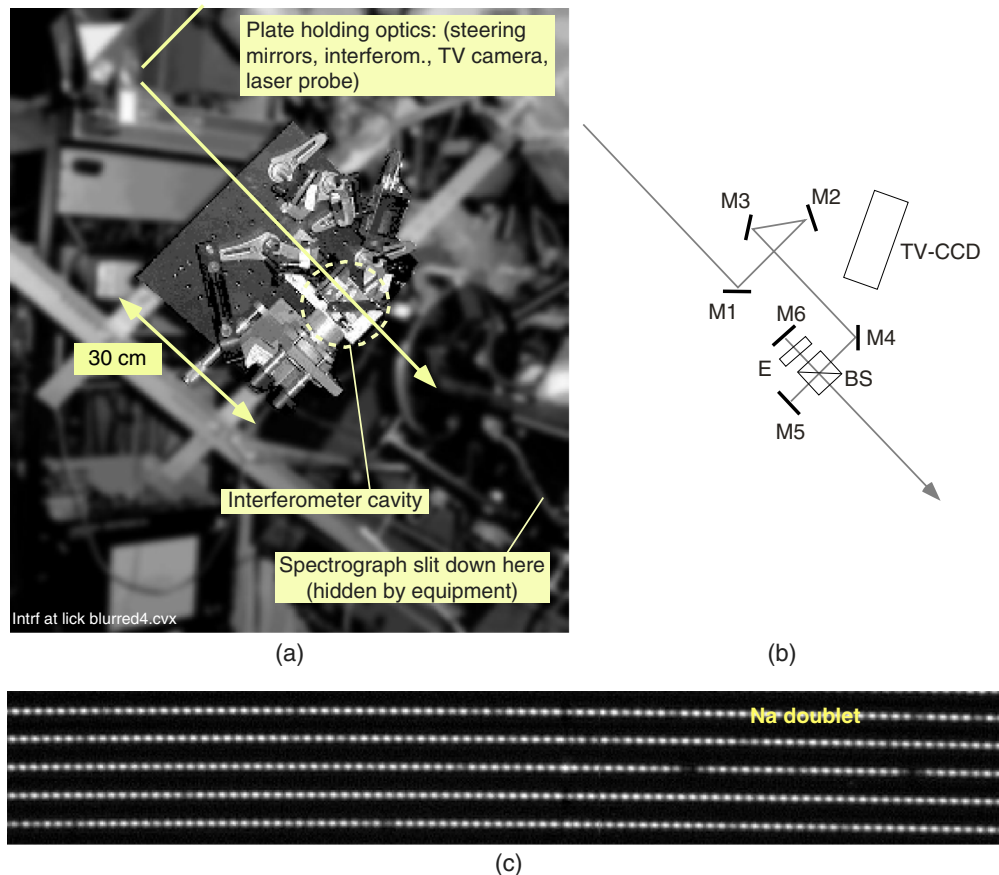
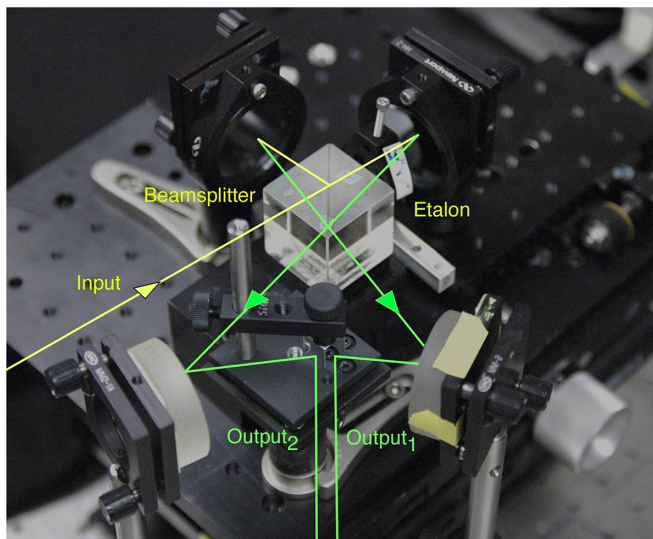


Fig. 10 In 2002, we tested a single-delay EDI on the Lick Observatory Hamilton echelle spectrograph, famous for exoplanet observations. We demonstrated that we can use a single-pixel wide beam [(c) shown with 1.1-cm delay] and do phase stepping versus time rather than spatially across a slit. Second, we can boost⁵ the spectral resolution of a facility spectrograph $\sim 2\times$ with a single delay of 3.3 cm, without narrowing the slit or changing its grating. The interferometer mounted on a 30-cm square plate (a) temporarily intercepts, interferes it with 1.1- or 3.3-cm delay, and then returns the beam from the telescope to the spectrograph entrance slit, without disturbing either system. (b) Schematic: an even number of reflections off mirrors M1 to M6 and BS makes alignment easy—insensitive to plate orientation. The TV monitors delay stepping phase via HeNe laser fringes through the same cavity, and M5 was moved via a PZT.

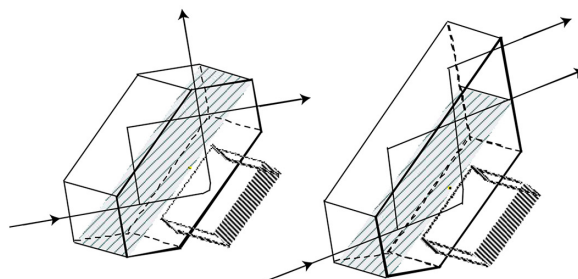
envelope. Then one can truly say that for features broader than the linewidth the EDI acts like a dispersive spectrograph (the wavelet waves average away), and for features narrower than the linewidth it acts like an interferometer (the wavelet waves dominate). But usually, we prefer to display just the fringing component in the EDI instrument response since that is what is new and different. How the combination of fringing and non-fringing terms behave with photon noise, or with detector noise, is described in Ref. 12.

The two techniques EDI and FTS share the need to have a synthetic output spectra (b), (f) that has finer bins than the detected spectra, because these two technique frequencies shift the detected spectra up to higher frequencies. The purely dispersive technique does not need a rebinned output spectra, and the detector spectra is often sufficient, but we perform rebinning of its detected spectra to facilitate comparison to the other spectra that have higher resolution features. In discussions about EDI, there is potential for confusion since both spectra types are sensible but hold different shapes. Moiré patterns (a) (phase stepped intensity changes) on the detector get frequency upshifted (multiplication by $e^{-i2\pi\nu}$) to wavelets (b) in the output spectrum [Eq. (27) of Ref. 11].

We rebin to a uniform 0.05 cm^{-1} spacing. We have used both Fourier zero padding interpolation (where the number of points of the FT is increased and the new high frequencies filled



(a)



(b)

Fig. 11 (a) An interferometer splits light into two complementary phase outputs, both of which can be directed to the spectrograph (but should fall on different pixels). This was done with the first version⁸ of T-EDI at Mt. Palomar (2007 to 2009). The interferometer is ideally lossless, but practical optics has parasitic reflections, which can be reduced by high-quality anti-reflection coatings. Because the optical delays needed are typically 1- to 5-cm round trip, produced by 0.5- to 2.5-cm single-trip distances, the interferometer can be a few cm in size. (b) Monolithic designs made from glued prisms would reduce vacuum-glass interface losses and be especially compact. One could envision a rotary “filter” wheel that held several small monolithic interferometers having different delays, so that instead of swapping just the glass etalon component of a stationary interferometer, the entire interferometer was swapped into the beam prior to the spectrograph. Monolithic graphics reproduced with permission from Ref. 7 by SPIE.

with zeros then inverse Fourier transformed) or cubic spline interpolation for rebinning. The cubic spline produces slightly less ringing in the wings. The T-EDI detector pixel spacing varies across the band and several echelle orders, $\sim 0.6 \text{ cm}^{-1}$ per pixel for 4000 to 5300 cm^{-1} order, to $\sim 1.2 \text{ cm}^{-1}$ per pixel for 8000 to 10,600 cm^{-1} order.

7 Crossfading Theory

In this discussion of fringing spectra, it is useful to visualize the phase plotted vertically, as if it was multi-phase data, even if data were taken in uni-phase mode.

7.1 Behavior in Dispersion-Space for a Drifted Spectrum

Features can be unblended (isolated) or blended where they are an unresolved group of more than one line. Examples of isolated solar lines are the left side of Fig. 8(a). Examples of blended solar lines are on the right side, where the moiré appears to have a slant. Blended iodine features

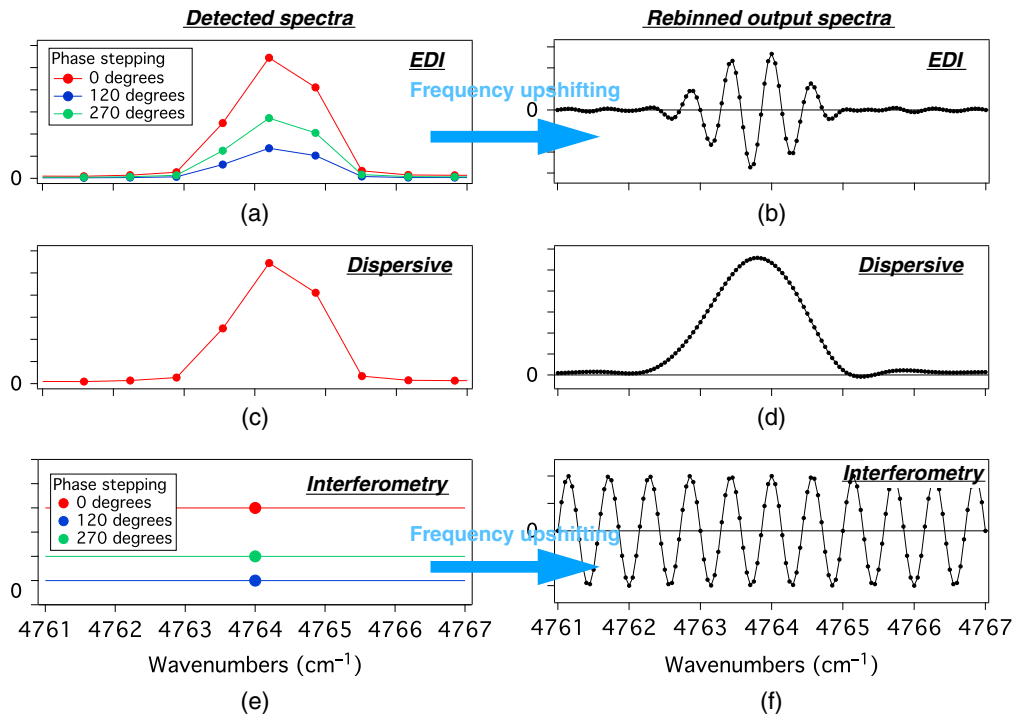


Fig. 12 Example instrument lineshapes for (a) T-EDI, (c) purely dispersive, and (e) purely interferometric (FTS). Both EDI and FTS require a synthetic output spectra rebinned to much finer bin spacing (0.05 cm^{-1}) than the detected spectra, to manifest high-resolution features. Pure-interferometry such as FTS has a single-phase stepped detecting pixel (e) collecting all wavenumbers and contributes a wide sinusoid $\cos 2\pi\nu\tau$ to the output (f), $\tau = 1.75 \text{ cm}$. EDI (a), (b) has hybrid behavior: inside the linewidth (c), (d) of the dispersive spectrograph it acts like the interferometer (e), (f). Thus from a phase stepped moiré (a) it generates a wavelet (b) with frequency set by delay τ . Vertical axis units are arbitrary. The phase steps here are notional.

form the smile-like moiré in Fig. 8(b). Figures 8(c) and 8(d) show how these smiles are formed from a group of nearly periodic lines whose periodicity is slowly changing.

7.1.1 For an isolated narrow feature

An isolated narrow spectral line is very much more robust to drift Δx for an EDI than for a native spectrograph alone. This is because the EDI periodic sinusoidal comb is embedded in the input spectrum and drifts along with it. The moiré phase only changes when the input spectrum moves relative to the sinusoidal comb.

7.1.2 For a blended feature

Blended features create slanted moiré, whose slope creates a phase error when translated by Δx . These errors vary randomly in polarity for different features, so it would tend to diminish (as square root of number of features) when summing Doppler velocity over a wide band. However, having them is still undesirable, because their irregular bipolar placement might couple to irregularities in the spectrograph having similar length scale, potentially creating a net error.

7.2 Behavior in Fourier Space for a Drifted Spectrum

7.2.1 Fourier behavior of conventional spectrograph

Figure 13 shows a plot of the modulation transfer function (MTF) that shows the effect of Δx on the Fourier response of both (a) a conventional spectrograph and (b) an EDI. The frequency

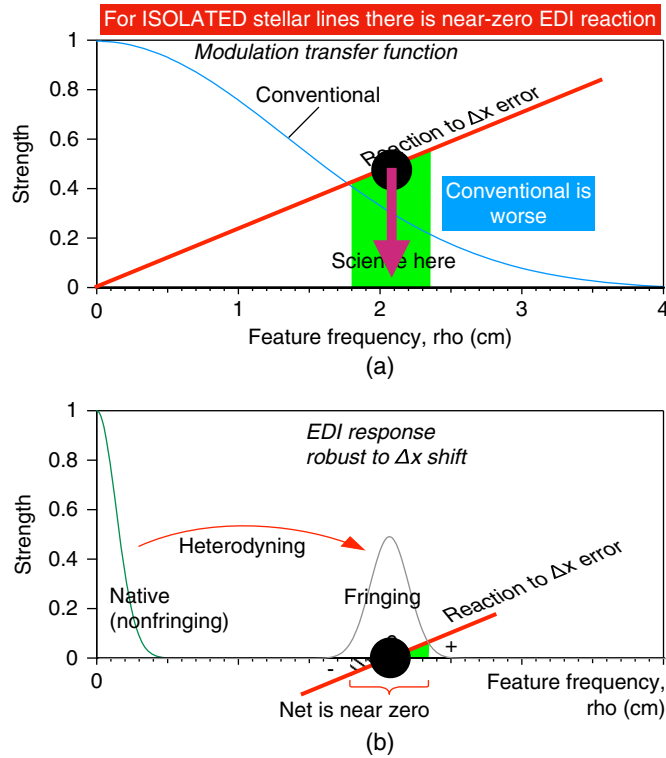


Fig. 13 Effect of a wavenumber (or wavelength) drift on the Fourier response, for (a) conventional and (b) EDI techniques, for the case of an isolated narrow feature (unblended). The reaction of the conventional spectrograph [black dot in (a)] is significantly larger than the near-zero reaction of EDI (b). An isolated very narrow feature produces a smooth distribution of frequencies in the region of the EDI sensitivity peak. These nearly cancel, producing near-zero TRC (translational reaction coefficient), i.e., very high-stability gain. Figure reproduced with permission from Ref. 11 by SPIE.

variable is ρ , in cm units. For the conventional spectrograph (a), a standard mathematical relation says that a shift Δx in the spectrum rotates its complex Fourier value by phasor $e^{i2\pi(\Delta x)\rho}$, where ρ is the frequency variable. For small Δx , the linear growth of the rotation angle of $2\pi(\Delta x)\rho$ radians is depicted by the red inclined line from the origin. This depicts angular reaction or imaginary component (since the magnitude changes slowly, as a cosine to that angle).

The green region represents high frequencies where the science (Doppler shifts or positional information of narrow features) typically reside. The black dot in the green region of (a) represents the large amount of reaction to a Δx for the conventional spectrograph to be compared with the near zero amount of the black dot for the EDI in (b).

7.2.2 Fourier behavior of EDI for isolated lines

For EDI in panel (b), the heterodyning shifts the sensitivity peak from the origin to higher frequency by delay τ . A spectral shift by Δx causes the FT of the EDI output spectrum to rotate by $e^{i2\pi(\Delta x)(\rho-\tau)}$ (Sec. 9.4 of Ref. 11). Thus the red EDI reaction line goes through zero height at the center frequency of the peak, at $\rho = \tau$, marked by the black dot in (b).

What does a feature look like that has all its frequencies at τ ? It would be an ultra-wide sinusoid that matched the period of the interferometer comb. It would generate a moiré that was perfectly horizontal, which therefore had zero slope, so under an Δx translation it would not change its phase. Hence the black dot in (b) at the center frequency $\rho = \tau$ has perfect robustness to drift Δx and $TRC = 0$.

However, features in general have a finite width and thus occupy a range of frequencies, not a single frequency. A perfectly narrow isolated line would generate an ultra-wide range of frequencies symmetric about τ . Since there are positive and negative reactions on either side of the peak about τ , these would sum to zero reaction.

However, for a non-zero width isolated line there still can be a small reaction to Δx . This is because it does not have a perfectly flat distribution of frequencies. Instead, it has an approximate Gaussian distribution with more energy to the low-frequency side of the EDI sensitivity peak than the high side. Especially for a feature intrinsic width that is similar to interferometer comb spacing (which puts the shoulder of the Gaussian in frequency space where it is most rapidly changing at the EDI fringing peak frequency).

7.2.3 Fourier behavior of EDI for blended lines

For blended features, the frequency distribution in general can be lumpy and non-uniform about τ and thus could generate a small net reaction error. The worst case would be a concentration of energy in the frequencies just to one side of the peak and not the other. (This is similar to the case of an unbalanced insult examined below in Sec. 7.4.2.) Fortunately, the crossfading technique dramatically zeros out the reaction for all features, blended or unblended.

7.3 Crossfading Cancels Net Reaction

Since the precision demanded of measuring an Earth-like exoplanet (10 cm/s amplitude) or measuring a few cm/s cosmic red shift over a few years is so severe, even when using a calibrant such as an iodine cell there could be significant residual instrumental errors, as discussed in Ref. 1. Hence we seek to cancel or dramatically reduce any potential error due to drift Δx , in spite of use of simultaneous calibrants, by the method of crossfading.

The point of crossfading is to cancel the net reaction so that TRC is as small as possible, for all types of features. Figure 14 shows that by combining the reactions of multiple delays, rather than single delay, the net reaction could be reduced to zero over a range of frequencies τ_{-1} to τ_1 . The reaction line for an individual delay flips polarity at its delay value. Hence two overlapping delay peaks will tend to cancel their net reaction when the high- and low-frequency sides combine in a weighted sum during data processing. By choosing weights, we can ensure this cancellation. Figure 15(a) shows the frequency region in between two delays, and (b)–(d) how the complex value of the net frequency response, represented by two vectors \mathbf{V}_1 and \mathbf{V}_2 , sum to form a small angle, which can be zero when (d) the weights are adjusted.

7.4 Requirement of Pairwise Same Insult

A condition of ideal crossfading is that the same Δx applies to both delays (1 and 2) of a pair, so that $\Delta x_1 = \Delta x_2$, and that weightings w_1 and w_2 calculated under that assumption produce cancellation. If the Δx slowly changes with time, only the portion equal on both sides of the pair will

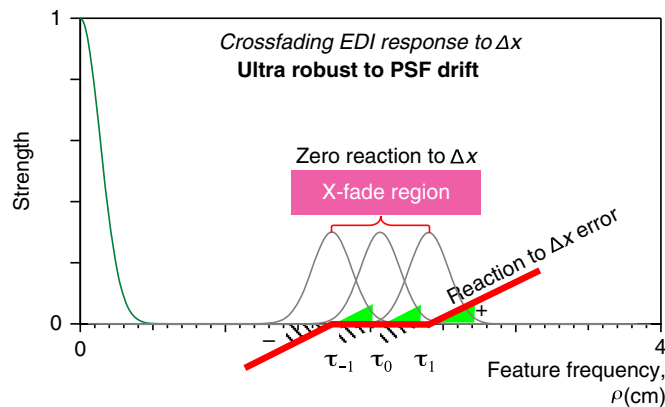


Fig. 14 With multiple delays (three shown), rather than single-delay EDI, the net reaction to an insult Δx is dramatically reduced to near zero over a range (between delay centers) when the positive and negative portions of the peaks overlap and appropriate weightings are used. This is called “crossfading.” Figure reproduced with permission from Ref. 11 by SPIE.

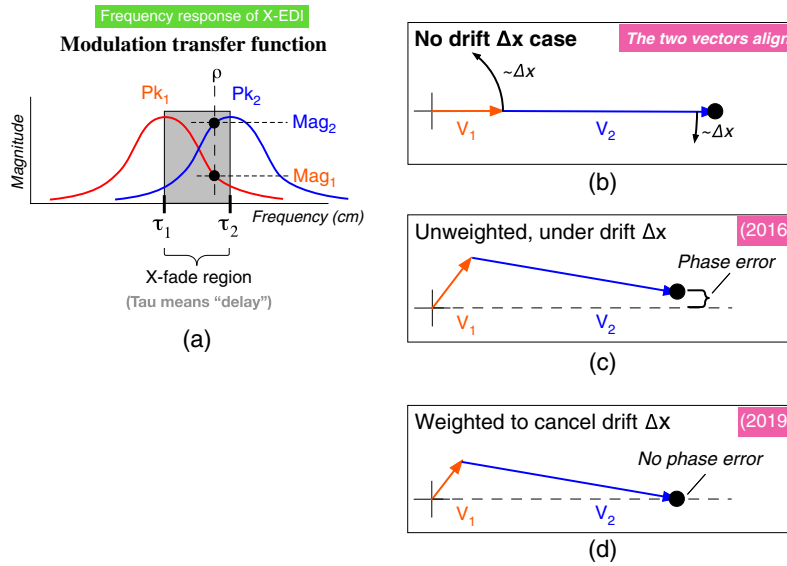


Fig. 15 (a) Two overlapping sensitivity peaks in the MTF of the EDI, one for each delay at τ_1 and τ_2 . These twist in opposite directions in response to a drift Δx and can be made to cancel if weights are chosen appropriately. (b) Virgin position for $\Delta x = 0$, depicted as zero for clarity but in reality can have a small angle that is removed during equalization. (c), (d) The vector sum of V_1 and V_2 have lengths Mag_1 and Mag_2 set by the peak height at ρ , and angles by Δx proportional to $(\rho - \tau_n)$. (c) The unweighted sum (as in the 2016 result Fig. 25) shows a reduction, but not perfect cancellation. (d) After weighting by w_1 and w_2 , the sum vector is zero (as in the 2019 result Fig. 26). This only works when the two components rotate in opposite directions to Δx , which only occurs when ρ is between τ_1 and τ_2 . (In these figures, delay and frequency have the same units of cm and the horizontal axis is interchangeably labeled.)

cancel. So instead of 0, TRC for a partially unbalanced case $\Delta x_1 - \Delta x_2 = \delta x$ could be as large as $\sim TRC_u(\delta x/\Delta x)$, where TRC_u is the 100% unbalanced case where $\Delta x_1 \neq 0$ and $\Delta x_2 = 0$, estimated below.

7.4.1 Slow time-dependent drifts

Drifts that change slower than the phase stepped exposures for a pair of delays are a candidate for crossfading. (Alternatively, if Δx grows in a predictable manner, such as a slow and linear thermally induced drift, one could calculate weightings anticipating that.)

Interestingly, only the time needed to complete an individual pair of delays, not all delays of a large multiple delay set, is what matters. Hence if readout noise was not an issue, one could optimize the exposure schedule to take alternating exposures of a first pair of delays 1, 2, 1, 2, etc., and then the next pair 2, 3, 2, 3, etc., which minimizes the unbalanced portion of insult δx in a pairwise fashion. This result holds provided the insult Δx is small enough to be in a linear reaction regime (which is true for most precision astronomical Doppler measurements, already using conventional stability mitigations) so that adding the same offset to both of the next pair Δx_2 and Δx_3 does not change their net subtracted reaction.

7.4.2 Estimation of effect of unbalanced insult on TRC

Figure 16 shows geometry to estimate $TRC_u \equiv TRC_{edi}$ for a 100% unbalanced insult, where $\Delta x_1 \neq 0$ but the second delay of the pair has unexpectedly $\Delta x_2 = 0$. This is the same as calculating the weighted TRC_{edi} to a single delay 1 in between τ_1 and τ_2 . (a) We use the phase reaction line for a conventional spectrograph (thin black), having $TRC_{conv} = 1$ by definition, to establish the value for TRC_{edi} (dashed). The shorter reaction line (bold) originating from τ_1 has the same slope of unity. Thus the ratio of heights in the middle of the overlap range (red dot height divided by green dot height) is given by $(\tau_2 - \tau_1)/(\tau_2 + \tau_1)$.

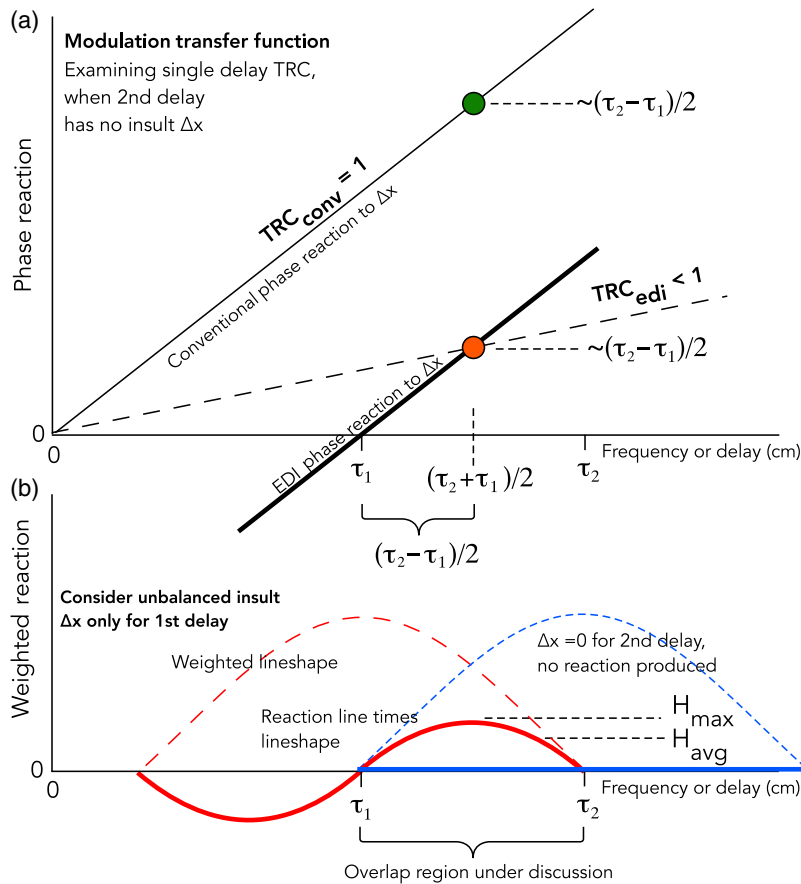


Fig. 16 Geometry to estimate TRC for a 100% unbalanced insult, where the second delay of a pair has zero insult, which is to calculate the weighted TRC to a single delay 1 in between τ_1 and τ_2 . (a) We use the reaction line for a conventional spectrograph, where $TRC_{conv} = 1$ by definition, to establish the value for TRC_{edi} . Since the phase reaction lines have the same slope, the ratio of heights in the middle of the overlap range (red dot height divided by green dot height) is given by $(\tau_2 - \tau_1)/(\tau_2 + \tau_1)$. (b) This TRC_{edi} should be reduced by multiplication by $(2/3)$ or $(2/\pi)$ since our overlap region of interest includes sections at the edges at τ_1 and τ_2 that have nearly zero reaction. The shape of the reaction region is an arch (red curve) found by multiplying the crossfaded weighted lineshape (dashed peak) by the sloping reaction line. This creates a sinusoidal-like shape having positive and negative lobes on either side of the delay center. If the weighted lineshape has triangular or sinc function shape, the lobe has parabolic or half-period sinusoidal shape, and the average height of those shapes is $2/3$ and $2/\pi$, respectively. If the spectral feature is narrow and isolated, then its frequencies would probably fill both positive and negative lobes and nearly cancel out, and its TRC_{edi} would be near zero. The worse case scenario is if we have a blended feature that has a majority of frequencies in between τ_1 and τ_2 , then it could have $TRC_{edi} \sim (0.63)(\tau_2 - \tau_1)/(\tau_2 + \tau_1)$. Plugging in values from the T-EDI example delay pair E4 and E5 of 0.96 and 1.27 cm discussed later, this produces $TRC_{edi} \sim 0.1$. Hence we would need a maximum 1% non-uniformity between insults Δx applied to the two delays, which could produce a random TRC_{edi} component of ~ 0.001 , comparable to the TRC expected to be produced by the T-EDI simulation.

In (b), the phase reaction lines are multiplied by the weighted lineshape (dashed peak), which describes signal strength at each frequency, to form an arch like shape (red curve) describing the final reaction. It is a characteristic of crossfading that the optimal weightings will force the final lineshape to be zero at the center of the other delay of the pair so that both reactions are always zero there, even in the normal balanced case where $\Delta x_2 = \Delta x_1$.

We do not know where the strongest frequencies of the input spectrum will lie, because we are being generic. If they lay symmetrically around τ_1 , the net reaction would be near zero and TRC_u near zero. Let us suppose a worse case scenario, but simple mathematically, that they

evenly fill the region τ_1 to τ_2 . Then the answer would be the average height of the arch, which is ~ 0.63 , since it is $(2/3)$ or $(2/\pi)$ for a triangle or sinc function weighted lineshapes (which are two convenient solutions that theoretically satisfy crossfading for small phase shift angles¹¹) that after multiplication by the phase reaction line create arches of parabolic or sinusoidal shape (red curve). The blue thick line depicts the zeroed out reaction of delay 2, which normally would have the same arch shape as the red curve of delay 1 but opposite in polarity so that it would exactly cancel.

Hence, we estimate

$$\text{TRC}_u \geq 0 \quad \text{and} \quad \text{TRC}_u < \sim (0.6)(\tau_2 - \tau_1)/(\tau_2 + \tau_1), \quad (2)$$

depending on where the frequencies of the source spectrum lie. For example, using values from the T-EDI example for delay pair *E4* and *E5* of 0.96 and 1.27 cm discussed later, this produces $\text{TRC}_u = \sim 0.1$. Since that is the value for a 100% unbalance in insult, the value for a smaller imbalance δx is $\text{TRC}_{\text{edi}} = \text{TRC}_u(\delta x/\Delta x)$ or $\sim 0.1(\delta x/\Delta x)$. So a 10% insult non-uniformity could produce TRC_{edi} between 0 and ~ 0.01 . Hence to measure a 0.001 TRC, similar to what we get from the T-EDI simulations, we would need a maximum 1% insult non-uniformity between the two delays *E4* and *E5*.

7.4.3 Fringing peak crossfaded with native peak

In many single-delay EDI,^{3-5,27} the fringing peak at least partially overlaps the native spectrograph sensitivity peak. (This is indicated in raw data by a partially resolved interferometer comb not completely blurred away.) This has the interesting property that Δx will automatically be the same for the native and the EDI peak—and thus could crossfade for drifts of all time scales. This is because the non-fringing (ordinary) spectrum and fringing complex spectrum are both manifested simultaneously in the same fringing spectrum data and separated during phase stepping data analysis.

The caveat is that the crossfading region is limited to frequencies from 0 to up to the center of the delay peak τ not higher. However, once the crossfading has been performed, the amount of actual Δx occurring in the data may be sensed by comparing the crossfaded and original spectra (after filtering away frequencies higher than τ). Then the frequencies higher than τ could be corrected using this measured Δx by conventional translation and combined with the lower frequencies to form a finished output spectrum. This idea is preliminary.

Second, the TRC's generated by a single delay low enough to overlap the native will not be as impressively small as for the case of much higher delays well out of overlap with the native peak. This is because TRC in Fourier space is a slope of phase error divided by delay, so higher delays tend to reduce TRC.

7.4.4 Pixel misplacement as time-independent drift

The detector pixel misplacement error type is time independent, and thus ideally suited for crossfading removal by a time sequential multiple-delay EDI (like T-EDI^{9,11,12}). This would use the same rows on the detector for each delay while changing among different delays—thus the same Δx pattern applies for each delay pair.

7.5 Staircase Mirror for Simultaneous Multi-Delay EDI Testbed

Figure 17 shows diagram of stepped mirror scheme^{2,11} for measuring multiple delays simultaneously that we propose for exploring the capabilities of crossfading experimentally. The at least two delays are imaged to different locations along the spectrograph slit. This allows error of all time scales, including fast drifts due to air convection or vibration, to be compensated by the crossfading method.

This is similar to scheme of 1998 constructed apparatus Fig. 7, which uses uncollimated beams imaged by lenses L_1 , L_2 external to the interferometer cavity, but with a staircase etalon and mirror, and dispersion direction into the paper instead of within page plane. Lens L_1 images

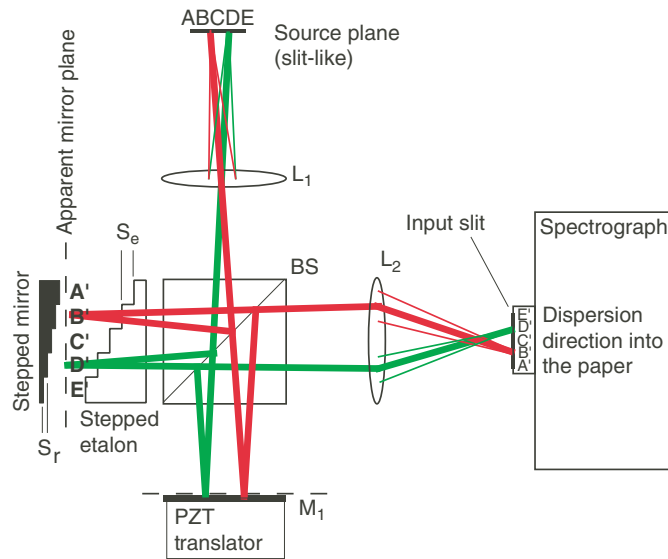


Fig. 17 Proposed scheme for a stair-stepped mirror interferometer^{2,11} for measuring multiple delays simultaneously with the same dispersive spectrograph, so that short-term spectrograph drifts Δx can be shared among at least two delays, to cancel drift using crossfaded data mixing. (This is lens and uncollimated beam imaging scheme of 1998 constructed apparatus Fig. 7, but with a staircase etalon and mirror, and dispersion direction into the paper instead of within page plane.) Lens L_1 images source slit to apparent mirror plane (dashed lines), and L_2 reimages that plane to the spectrograph slit (dispersion direction into the page). BS superimposes the apparent mirror plane with mirror M_1 , (complementary BS ignored). Slit-like source ABCDE illuminates step mirror segments A' through E' with a different delay for each step. Glass etalon step size S_e with refractive index n , and mirror step S_r are optimally related by $S_r = S_e(n - 1)/n$, to keep virtual image of mirror steps in apparent mirror plane, so that delay is independent of ray angle and has improved fringe visibility for extended sources (wide angle Michelson interferometer²⁸).

source slit to apparent mirror plane (dashed lines), and L_2 reimages that plane to the spectrograph slit (dispersion direction into the page). BS superimposes the apparent mirror plane with mirror M_1 , (complementary BS ignored). Slit-like source ABCDE illuminates step mirror segments A' through E' with a different delay for each step. Glass etalon step size S_e with refractive index n , and mirror step S_r are optimally related by

$$S_r = S_e(n - 1)/n, \quad (3)$$

to keep virtual image of mirror steps in apparent mirror plane, so that delay is independent of ray angle and has improved fringe visibility for extended sources [wide angle Michelson interferometer (WAMI)²⁸].

The extra time of flight from mirror step to the apparent plane is proportional to S_r . The extra time due to the slower propagation of light through glass is $S_e(n - 1)$. Combining those and doubling for the round trip gives us the interferometer delay in distance units:

$$\tau = 2S_e(1 - 1/n) + 2S_e(n - 1) = 2S_e(n - 1/n). \quad (4)$$

For example, the 6.35-mm-thick BK7 etalon ($n = 1.52$) used in Ref. 4, when used with the mirrors superimposed in the optimal WAMI position has $(2)(6.25)(1.52 - 1/1.52) = 10.95$ mm of delay. (One can deviate from the WAMI condition to slightly change the delay, and if the beams are collimated there is only minor loss of visibility.)

The precise delay achieved is found later spectroscopically, such as by measuring the white light generated interferometer comb $1 + \cos[2\pi\nu\tau(\nu)]$ and plotting fringe count versus wavenumber. The derivative will yield the delay $\tau(\nu)$, which will be a slow function of wavenumber because of glass dispersion in its index $n(\nu)$.

With the mirror M1 aligned normal to the beam the interferometer phase is uniform along the slit (the Y direction) and this is the uni-phase mode presenting the data [Fig. 33(b)]. Alternatively, the mirror M1 can be tilted slightly so that there is an extra delay that grows linearly along the spectrograph—this makes the multi-phase mode, which has a slanted interferometer comb [Fig. 33(a)].

If the staircase stepped mirror has facets that are not perfectly normal to the beam or to each other, that would slightly change the periodicity along the Y direction, which could also manifest as a phase offset. This is not a problem since the phase stepping analysis is designed to remove common mode behaviors such as this that affect many wavenumber (X direction) channels the same way. Consider that if a mirror segment has more than normal tilt, it applies that the same tilt to every wavenumber X channel; while stellar spectral features are changing rapidly with X , changing both the phase and magnitude of the fringe versus X . It is straightforward to solve for those instrumental artifacts that are constant or change very slowly from channel to channel. Also one can compare behaviors between stellar and calibrant signals taken either on the same pixels or nearby pixels and thus affected the same by the extra tilt of the mirror segment.

We are also not concerned about the different segments having slightly different tilts (which can affect phase steps) because the phase stepping analysis algorithm analyzes each mirror segment independently and solves for the precise value of the phase steps—we only tell it approximate values. Due to the massively parallel nature of fringing spectra (they are all affected by the same step in delay), a unique solution for the precise phase step values results from the sinusoidal fitting process.

To be concrete in discussion, we suggest an expedient way to construct a staircase mirror is to stack in a staircase arrangement several glass plates similar in size to microscope slides, which are ~ 1 -mm thick. (Of course, in practice would be better to use clearer and flatter optical quality glass such as BK7, instead of actual microscope slides.) The mirror staircase could be solid material machined to have steps of $(1.5 - 1)/1.5 = 0.5$ mm. One could glue first surface mirrors on these. A 1-mm thickness with an assumed index of 1.5 would produce a delay change of $2(1.5 - 1/1.5) = 1.66$ mm or 0.16 cm, when in WAMI configuration [found by maximizing the fringe visibility as the M1 (to BS) distance is changed]. Supposing we are in visible wavelengths of green ($0.5 \mu\text{m}$) (so we can use the iodine cell and green line of Hg), then there are $1.66/0.0005 = 3333$ wavelengths that fit inside the delay. Hence that would suggest a relevant resolution for the native spectrograph of about the same, at least 3333, to have some overlap between delays. Having higher resolution is useful practically, because one can always widen a slit to reduce resolution but not the other way round, and it also is better to have too much overlap instead of too little overlap.

So we suppose that we could again use the same $R = 20,000$ Jobin-Yvon spectrograph and use wider slits to reduce its resolution when desired. We can insert a normal 1.1-cm etalon (6.35-mm thick BK7) delay that fills the beam and increases all the separate delays, so they have values $1.1 + 0.16$, $1.1 + 0.32$, etc. in cm. This will make the device have a reasonable sensitivity to Doppler shifts and decrease the relative size of the delay step size (0.16) to the bulk delay (1.26 cm) so that worst case unbalanced TRC_u is small at $(0.6)(0.16)/(1.26) = 0.08$.

7.5.1 Calibrating away pixel misplacement

The delay channels are necessarily on different rows and so they may have a different error function Δx of “baked in” pixel misregistration for each delay. Crossfading will defeat this baked-in Δx insult, but we still want to measure this component quantitatively. This could be done by the following procedure where we strategically offset all the delays so that it effectively changes which rows are responsible for producing a fringing spectrum having a given delay. Then differences with an earlier fringing spectrum for that delay could only be due to the pixel misregistration.

Suppose we have three delays τ_1 , τ_2 , τ_3 , of the same interval $\delta\tau = \tau_2 - \tau_1 = \tau_3 - \tau_2$. After taking data for the three delays normally, we suggest inserting into the same staircase arm an additional etalon of the staircase step height (1 mm) but with the usual wide width that fills the beam and all staircase segments. (And then mirror M₁ is shifted about half that amount to maintain the WAMI condition.) This collective increase in delay causes τ_1 to effectively becomes τ_2

but still using the misregistrations of row #1. Then the opposite can be done. Removing that wide etalon and placing it in the other arm causes a decrease in delay by $\delta\tau$, and thus τ_2 delay becomes τ_1 and τ_3 delay becomes τ_2 . By having several means for achieving the same delay but with different rows of baked-in pixel misplacement error function, we can isolate this error component.

7.6 Changing Overlap Across the Band

Figure 18 shows the degree of overlap of two EDI peaks at delay 1 and delay 2, or τ_1 and τ_2 decreases versus ν across a wide bandwidth. The EDI peaks are the same width and shape as the native spectrograph peak centered at zero frequency, just shifted up by τ and halved in amplitude. Grating spectrographs typically have constant resolving power $R_{conv} \sim \nu/\delta\nu \sim \lambda/\delta\lambda$ set by grating groove density. Since the resolution element $\delta\nu$ increases as ν/R_{conv} and from the uncertainty principle $\delta\rho \sim 1/\delta\nu$, then $\delta\rho \sim R_{conv}/\nu$. The positions of the peaks τ_1, τ_2 are fixed, set by the optical path length difference between two interferometer arms.

This causes the optimal weightings for crossfading to change across the band. At some high ν , there would be insufficient overlap. (If the weighted peaks were triangular-like, the corner of one peak should be located at the center of the neighboring peak.) One should select the delays and/or spectrograph resolution for sufficient overlap when the peaks are narrowest (high ν) since there is no penalty for excess overlap.

7.6.1 Boosted resolving power increases with wavenumber and max delay

By concatenating several delay peaks together in a cross-fading manner, shoulder to shoulder, a multiple-delay EDI effectively has a rectangular sensitivity curve from the minimum delay or the native spectrograph peak (around zero) up to some maximum delay τ_{max} . From this, we can produce an effective resolution element of approximately $\delta\nu \sim 1/\tau_{max}$, producing a resolving power $R_{edi} \sim \nu/\delta\nu \sim \nu/\tau_{max}$ that increases proportional to wavenumber, rather than being constant typical of grating spectrographs, and increases proportional to maximum contiguous delay τ_{max} .

After crossfading has concatenated the peaks together, then during the subsequent equalization step, besides dividing out the minor dips between peaks, one imparts an overall truncated Gaussian shape. Ideally, this approximates a true Gaussian having resolution R_{goal} . But because the Gaussian must be truncated at the high-frequency end $\rho = \tau_{max}$, there is ringing produced.

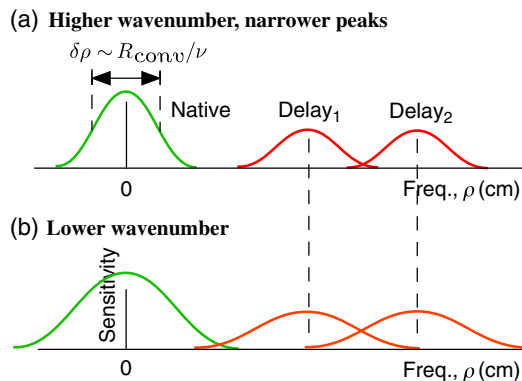


Fig. 18 The width of the sensitivity peaks is narrower at (a) high wavenumber than (b) low. The native disperser peak (green) sets the widths of all delay peaks (red). Grating spectrographs typically have constant resolving power $R_{conv} \sim \nu/\delta\nu \sim \lambda/\delta\lambda$ set by grating groove density. Since the resolution element $\delta\nu$ increases as ν/R_{conv} , and from the uncertainty principle $\delta\rho \sim 1/\delta\nu$, then $\delta\rho \sim R_{conv}/\nu$. The positions of the peaks τ_1, τ_2 is fixed, set by the optical path length difference between two interferometer arms. The optimal weightings for crossfading hence can change across a wide band. One should set the delays for sufficient overlap when the peaks are narrowest (high ν) since there is no penalty for excess overlap.

The amount of ringing energy is related to the difference in area between the true Gaussian and the truncated Gaussian. Hence, there is a trade-off between final Gaussian resolution and ringing in the net PSF.

An FT spectrometer has the same issue since by scanning a delay (having extremely narrow peak) continuously over a range up to τ_{\max} it also assembles a rectangular area of information about the FT of the input spectrum, which it customarily needs to impart a Gaussian shape. The latter is because a PSF that lacks large ringing is less confusing for distinguishing two neighboring lines.

However, some types of spectroscopy, such as measuring Doppler shifts, do not require a Gaussian PSF (provided the PSF is the same for both dates of measurement). If so, the higher frequency information sensed by the high delays can contribute during analysis instead of being apodized away. And high delays are more sensitive to a given Doppler shift, their reacting phase shift being proportional to τ .

8 T-EDI Project Testing Multiple Delays

Figure 19 shows a photograph and schematic of the T-EDI apparatus, which was the first multiple delay EDI to be tested on starlight and achieved 10x resolution boosting. (A previous multi-delay EDI was tested from 2003 to 2004 in the laboratory on the iodine spectrum and achieved 6x resolution boost.⁶) The interferometer assembly (black and silver) sits on top of TripleSpec NIR spectrograph²⁹ (blue cylinder) into 1-m Cassegrain cavity of 5-m Hale telescope mirror. The first (2007 to 2009) version used both complementary outputs. The second version (2010 to 2011) used one of the outputs as an input for calibration line and changed the beam imaging to collimated beams intracavity to improve the phase overlap between the starlight and calibrant.

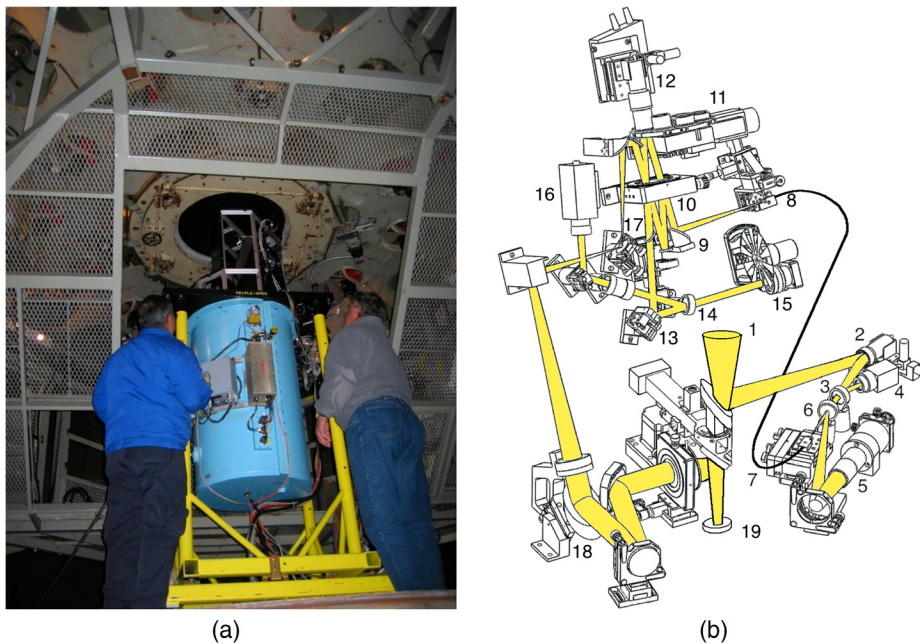


Fig. 19 (a) The T-EDI interferometer (black and silver) sits atop NIR TripleSpec dispersive spectrograph²⁹ (blue cryogenic cylinder), which bolts to Cassegrain output (39-in. diameter cavity) of the Mt. Palomar 200-in. mirror. (b) TEDI 2.0⁹ intercepts starlight beam (1) heading for TripleSpec and uses a dichroic split input tracking system (3) to mixing in ThAr lamp (5) to feed a symmetric, collimated interferometry cavity (BS 10). One interferometer output (13) is dichroic split (14) to a sensitive chopped IR diode (15) for flux maximization and a fringe tracking camera (16) for cavity nulling. The complementary interferometer output (17) is open-path relayed (18, 19) to the Triplespec slit. Photo reproduced with permission from Ref. 11 by SPIE. Schematic reproduced with permission from Ref. 9. © The Astronomical Society of the Pacific. All rights reserved.

8.1 Processing Raw Exposures into Complex Fringing Spectra

Figure 20 shows subset of raw T-EDI data at different interferometer phases, for two fibers A and B along one of several orders (A through E) in the NIR. Fiber A is ThAr lamp-alone and B is ThAr with starlight. As the delay is commanded to step by $0.25 \mu\text{m}$ increments, this creates a 0.19 cycle phase step, at the $\sim 7525 \text{ cm}^{-1}$. The intensity of a given feature is sinusoidally with phase. This more easily seen in Fig. 21 where lineouts for a given fiber along the (slightly curved) order are stacked so that phase varies vertically. The sinusoidal variations of the narrow ThAr lines appear as vertical dashes. The stellar fringes are much weaker (just a few percent) and appear as a smudge. In this raw form they are difficult to see by eye underneath, the stronger overall variations in intensity caused by clouds and seeing, which will eventually be divided out.

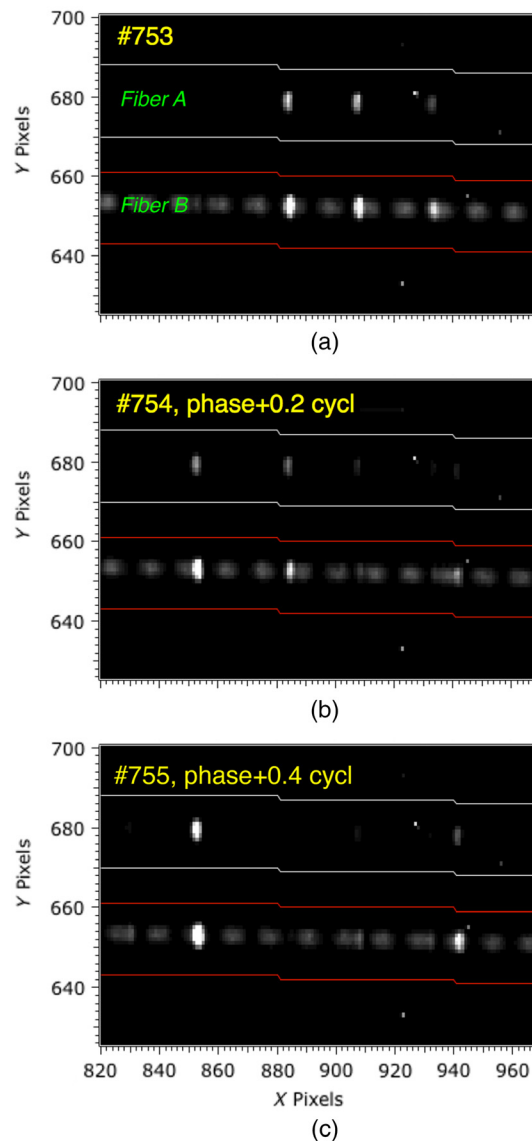


Fig. 20 (a)–(c) Three closeups of the TripleSpec²⁹ echelle C-order for ThAr lamp + star HD962, as interferometer delay changes for the E1 etalon ($\sim 0.1 \text{ cm}$). Fiber A is ThAr lamp, and fiber B is ThAr lamp with starlight. The $0.25\text{-}\mu\text{m}$ commanded delay change creates a 0.19 cycle phase step at $\sim 7525 \text{ cm}^{-1}$. Horizontal scaling $\sim 1 \text{ cm}^{-1}$ per pixel. Each feature’s intensity varies sinusoidally with interferometer phase, as seen when lineouts for a given fiber along the dispersion direction are stacked so that phase plots vertically (Fig. 21). Graphics reproduced with permission from Ref. 11 by SPIE.

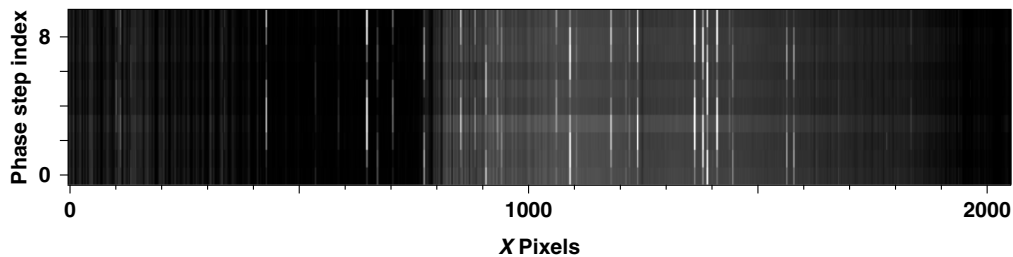


Fig. 21 Stacked set of 10 phase stepped exposures, for star GJ15A plus ThAr lamp, C-order of echelle. In this manner, modern uni-phase EDI echelle spectrograph data (~ 1 pixel high) appears as if in the multi-phase format of earlier EDI instruments using linear spectrographs, having phase sloping across a slit many pixels high. Horizontal pixels cover ~ 6742 to 8852 cm^{-1} . Y axis is file number (phase stepping index). Fringes versus Y are clearly seen in the bright narrow ThAr lamp lines. Starlight appears as smudge from $x = 800$ to 1500 . Its fringes are too faint (few percent) to be seen relative to the stronger seeing and variable cloud effects but appear mathematically in output.

Then essentially a sinusoidal fit is made along each column, and the sine and cosine amplitudes are assigned the imaginary and real parts of the complex fringing spectrum $\mathbf{W}(\nu)$, as illustrated in Fig. 22. The intensity offset of the fit is assigned the ordinary (native) spectrum $B(\nu)$.

8.2 T-EDI Delay Set

Figure 23 shows eight delays used in two instances, labeled $E1$ through $E8$ having default (b) values 0.083, 0.34, 0.66, 0.96, 1.27, 1.75, 2.92, and 4.63 cm. An eight-position filter wheel held eight glass etalons that were rotated into the interferometer cavity during a time sequence of exposures. Panel (a) shows a ninth etalon $E6.5$ of 2.38 cm was used in a few measurements to enable much higher resolution (the $10\times$ boost demonstration in Fig. 24), swapping out the position of the lowest etalon $E1$ 0.083 cm, and filling in the delay gap between $E6$ and $E7$ of 1.75 and 2.92 cm. Panel (b) shows the set used in the crossfading simulation using $E1$.

8.2.1 Extending the delay set for enormous R boosts

Figure 24 shows the dramatic $10\times$ resolution boost (native-only $R = 2700$ to EDI output of $R_{\text{goal}} = 27,000$) possible when the delay gap at 2.38 cm was filled with $E6.5$. (Only a portion of C-order processed spectrum shown.) Note that the native spectrograph (green dashes) cannot resolve many of the fine features. Consider that still there is a gap between $E7$ and $E8$. If additional etalons were used (using a 10-position rotary filter holder), the resolution R_{goal} could be extended further to $27,000(4.63/2.92) = 43,000$, (where we multiply by ratio of τ_{max} , $E8/E7$). By placing two rotary holders in series, we could have two delays (fine and coarse) sum—so using a 5-cm second delay with $E1$ to $E8$ shifts them effectively into the 5- to 10-cm range. Extending τ_{max} to 10 cm would produce resolution of $27,000(10/2.92) = 93,000$, a boost of $34\times$.

Significantly, the additional delays would for the first time create overlap with high delay $E8$ and permit crossfading to be performed at a high delays where Doppler velocimetry produces a larger signal (proportional to τ , provided the stellar spectrum is not rotational blurred and thus has energy at high frequencies). The primary purpose⁹ of the T-EDI was to measure Doppler velocity of M-stars, and we only had isolated delays $E7$ and $E8$.

8.2.2 Choosing the equalized output resolution

In EDI spectral reconstruction (SR), the final resolution R_{goal} is chosen by the user and has a shape in frequency space of a Gaussian (dashed curve). During equalization process, the bumps in the measured sensitivity of the combined EDI peaks are divided out so that their final strength has a smooth Gaussian, except in regions where there is no peak presence. Since a perfect

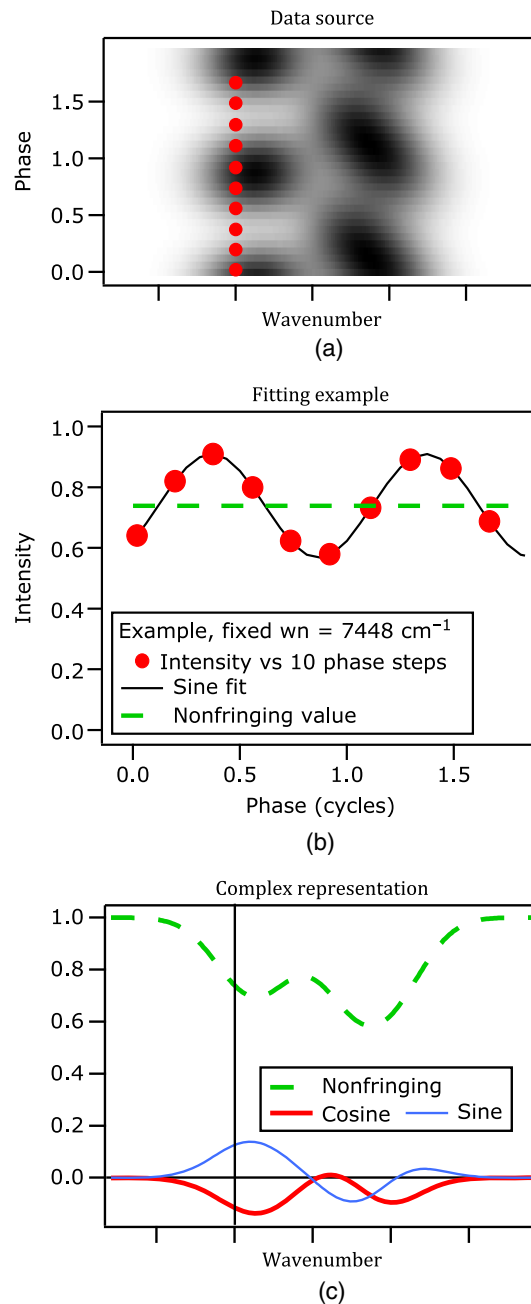


Fig. 22 Making non-fringing (native or ordinary) $[B(\nu)]$, and fringing spectra $[\mathbf{W}(\nu)]$ from stacked phase stepped spectra (a) (such as Fig. 21), essentially by fitting each column to a sinusoid (b) and assigning the cosine and sine components (c) to the real and imaginary parts of $\mathbf{W}(\nu)$. The intensity offset fitting result becomes $B(\nu)$. Graphics reproduced with permission from Ref. 12 by SPIE.

Gaussian is infinitely wide, at some point it must be cut off. (Technically, this creates ringing, although it can be negligible.)

Instance, Fig. 23(b) shows the delay set used for the crossfading demonstration, which was the default set not having $E6.5$, but $E1$ instead. Because we wanted to see how small of a TRC we could achieved, we desired maximum overlap between peaks, which occur for the lower delays $E1$ to $E5$. Hence we chose a conservative $R_{\text{goal}} = 9000$ and also chose the A-order, which has the lowest wavenumber and hence widest peaks (as seen in Fig. 18).

This R_{goal} has negligible presence beyond $E6$, so delays $E7$ and $E8$ do not effectively contribute to the sum, which forms the EDI output spectrum (red curves in Fig. 1). Second, they are

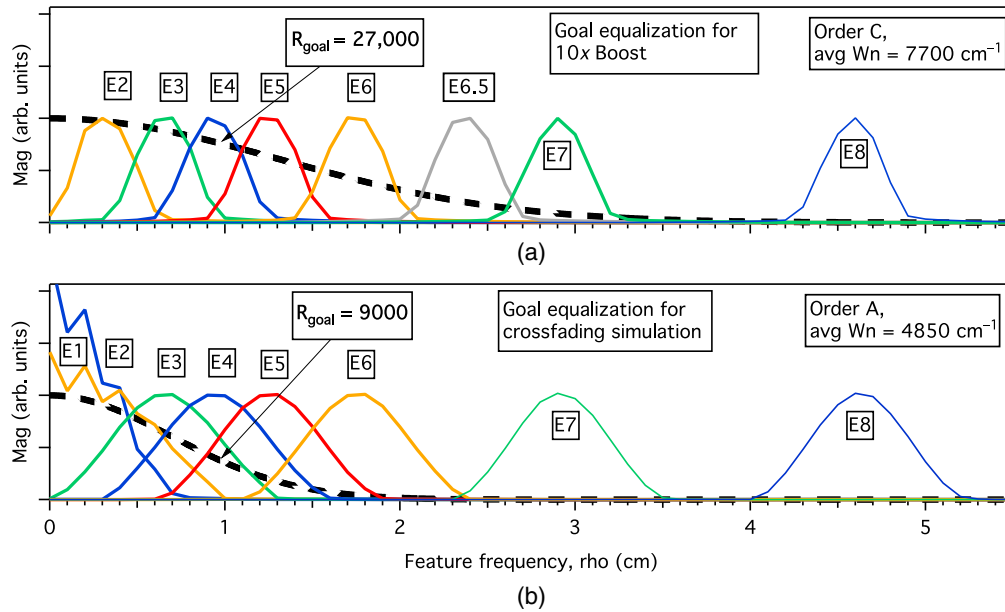


Fig. 23 Set of delays used for T-EDI, for two cases, shown in this MTF. (a) boosting resolution 10x in 2011 (see Fig. 24) and (b) simulating crossfading in 2019 (current report). Each delay creates a separate peak centered at the delay value, having shape of the native spectrograph response (not shown). Default eight delay values *E1* to *E8* were 0.083, 0.34, 0.66, 0.96, 1.27, 1.75, 2.92, 4.63 cm, respectively. (a) Using a series of contiguous delays (*E2* to *E6*, *E6.5*, and *E7*), a wide conglomerated MTF is formed. The 2.38-cm *E6.5* was swapped with the less important 0.08 cm *E1* delay in its filter wheel position to fill the gap between *E6* and *E7*, which would limit the resolution. After equalizing to produce an ideal Gaussian shape (red dash), an effective resolution of 27,000 can be achieved at ~7700 cm⁻¹. (b) For the crossfading simulation, we used a lower goal resolution of 9000 at ~4800 cm⁻¹ to benefit from the larger overlap between smaller etalons.

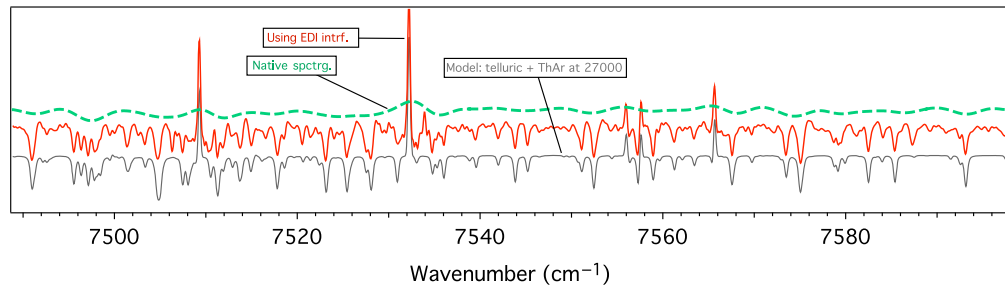


Fig. 24 Demonstration of a 10-fold resolution boost on T-EDI in 2011 observing telluric features mixed into spectrum of star κ CrB along with ThAr calibration lamp emission lines. The green dashed (top) curve is the ordinary spectrum measured without interference, having native $R = 2700$. It cannot resolve the telluric features. The red (middle) curve is the EDI (T-EDI) reconstructed spectrum measured with 7 contiguous delays, up to 3 cm, including the normally absent delay *E6.5* at 2.38 cm, and equalized to $R = 27,000$. The gray (bottom) curve is telluric model³⁰ and ThAr³¹ features blurred to $R = 27,000$, showing excellent agreement with EDI output. Graphics reproduced with permission from Ref. 11.

isolated and not overlapped, so we put their crossfading weights to zero to prevent them from contributing an unbalanced reaction. However, *E7* and *E8* still appear in the wavelet stack since this stack shows the wavelets prior to amplitude modification. (We find *E7* and *E8* useful to guide the eyes and also to suggest that the further precision one could achieve if those higher delays were used.)

8.3 Artificially Applied Insults for the Demonstration

To simulate crossfading on read data, but older data prior to crossfading invention, we artificially apply an insult Δx to all delay channels. This because we lack an experimental multiple-delay data set where we can be sure, the same drift was present for all delays. In 2011, at the time of the last multiple-delay data taking of the T-EDI project, we had not yet conceived of the crossfading technique and thus were unaware of the need to take data in pairs of delay in rapid succession. Instead, we took many exposures sitting on a single delay and then moved to the next delay and so on. There was significant drift between delays of order 0.4 cm^{-1} per hour (see Fig. 36 of Ref. 11). This drift was removed during analysis by conventional means, by comparing measured to theoretical³¹ ThAr lamp spectra, for the writing of the 2016 journal article on T-EDI.¹¹

Then in 2019 after crossfading was further developed, for the demonstrations of this report, we used that already aligned data set as a starting point and manually applied a Δx . We used our working code to Fourier process the moiré data, which reverses the heterodyning and sums the wavelets to form an output spectrum. It was our usual working software but with added code to apply a frequency-dependent weight prior to the summation.

8.4 Review: Prior Analysis Using Small Argument Approximation

Sections 4–7 in Ref. 11 describe our EDI algorithm for processing of echelle fringe spectra data, performing heterodyning reversal, and summing and equalizing the wavelets to form an output spectrum, and Sec. 10 describes some theory of crossfading. When a small shift Δx is observed in an EDI output spectrum for a given delay τ , its FT rotates in complex space as the phasor $e^{i2\pi\Delta x(\rho-\tau)}$, twisting about the center of the peak at $\rho = \tau$. The imaginary part of this represents the angle of the wavelet phase shift. This has sinusoidal behavior $\sin 2\pi\Delta x(\rho - \tau)$, which for small arguments is linear in $\Delta x(\rho - \tau)$. The analysis done in Ref. 11 assumed this small argument linearity and described theoretical lineshapes, such as the triangle or the sinc function, that exactly cancel the net phase reaction to an insult if both high- and low-delay overlapping peaks had these shapes.

This theoretical lineshape analysis mostly occurred in the detector space, prior to the reversal of heterodyning that shifts the low-frequency moiré pattern up to the original high frequency. This way, since all the detected moiré for each delay will be under the same lineshape, all the frequency up-shifted lineshapes will also be the same shape. So we searched for “magic” shapes such as the triangle or sinc function that would be desirable for the measurement of the moiré pattern in the detector wavenumber space (as opposed to the output wavenumber space, which is up-shifted).

From Ref. 11, we describe a family of curves that satisfy this condition of producing TRC = 0 in this crossfade region between the peaks. The net reaction to a horizontal PSF drift is the product of the line $(\rho - \tau_n)$ through the center of the peak times the LSP or PSF $\text{psf}(\rho)$ of each peak. For n peaks, this is

$$\text{reaction}(\rho) \propto \sum (\rho - \tau_n) \text{psf}(\rho - \tau_n). \quad (5)$$

To produce perfect cancellation between two overlapping peaks of the same shape $\text{psf}(\rho - \tau_1) \equiv f(r)$ and $\text{psf}(\rho - \tau_2) = f(r - 1)$, and using a dimensionless frequency $r = (\rho - \tau_1)/|\tau_2 - \tau_1|$ centered at the left peak #1, we must satisfy

$$rf(r) + (r - 1)f(r - 1) = 0. \quad (6)$$

A triangle satisfies this equation, with the left and right terms forming positive and negative parabolas that exactly cancel each other. Colleague Linder³² pointed out that a sinc function $f(r) = \sin(\pi r)/r$ also satisfies this equation, in which case instead of parabolas, we have a positive and negative sine peaks cancelling each other out. The common characteristic of the solutions to Eq. (6) is that they have sharp corners that cross zero linearly locally, and each corner is positioned exactly at the center of the neighboring peak so that when it is at maximum the other contribution is zero.

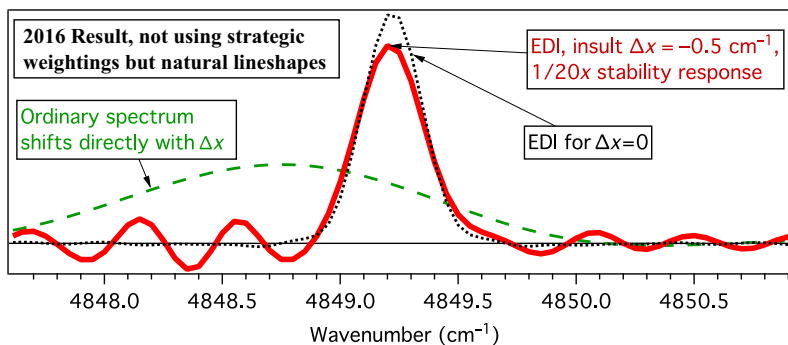


Fig. 25 Previous demonstration reported in 2016 (Fig. 42 of Ref. 11) using simple unweighted (natural lineshape) summation on multiple delay T-EDI data of ThAr spectral lamp with an artificial shift of $\Delta x = -0.5 \text{ cm}^{-1}$ applied to each delay's data. A 20 \times stability gain resulted since the EDI peak (red curve) shifted -0.025 cm^{-1} relative to the $\Delta x = 0$ case (black dots). In contrast, the conventional spectrum (green dashes) shifts directly with Δx .

8.5 Prior Crossfading Results Good but Imperfect

The concept of crossfading was introduced in Ref. 11 in 2016, but the calculation of idea weights was not fully implemented. However, even without weights to modify the lineshape, the overlapping that occurs with the unweighted natural lineshape of two neighboring delay signals provides some “accidental” crossfading [as vector diagram Fig. 15(c) suggests]. This produces a reduction, although not to zero, of TRC. An example is Fig. 25, which shows the 2016 good but imperfect result of the crossfading simulation (originally Fig. 42 of Ref. 11). This was a T-EDI measured ThAr line at 4849 cm^{-1} whose data were artificially shifted by $\Delta x = -0.5 \text{ cm}^{-1}$. Without using crossfading weights, a reaction of -0.025 cm^{-1} was observed—hence $\text{TRC} = 0.05$ (stability gain of 20 \times).

8.6 Improved Method for Calculating Crossfading Weights

We have now (2019) repeated the simulation on the same ThAr line but with an improved algorithm that uses crossfading weights, and we achieve $\text{TRC} \sim 0.001$ (stability gain of 1000 \times), as shown in Figs. 1 and 26. (Figure 25 was redrawn in the same wavenumber range and curve colors as Fig. 26 to facilitate comparison.) This is a 50 \times improvement over the 20 \times result of 2016, which is mainly due to the use of explicitly calculated weights to produce nearly perfect vector cancellation [as in Fig. 15(d)]. Note there is also a lot less ringing than in Fig. 25, which indicates an ideal Gaussian shape in frequency space is maintained despite insult Δx .

The net EDI (red curve) is a sum of weighted wavelets, each having different delays. The weights were calculated from the behavior under Δx of a neighboring 4764 cm^{-1} line acting as ersatz calibrant then applied to the 4849-cm^{-1} subject line. About 0.001 cm^{-1} peak shift was observed comparing $\Delta x = -0.5$ to $+0.5 \text{ cm}^{-1}$, for $\text{TRC} = 0.001$. The black dotted curve is $\Delta x = 0$ case. The ordinary spectrum (green dashes) responds directly to Δx .

8.7 New to the Algorithm Since 2016

Since the introduction of the crossfading concept in 2016 (Sec. 10 in Ref. 11), we have made algorithmic improvements:

8.7.1 Use smaller region of interest

We now use of a minimal number of points (200 at 0.05 cm^{-1} per point or 10 cm^{-1}) for the region of interest around the feature, instead of full A-order size of 28,000 points. (All orders of the raw T-EDI data, 2048 pixels wide, are resampled to a uniform 0.05 cm^{-1} spacing.)

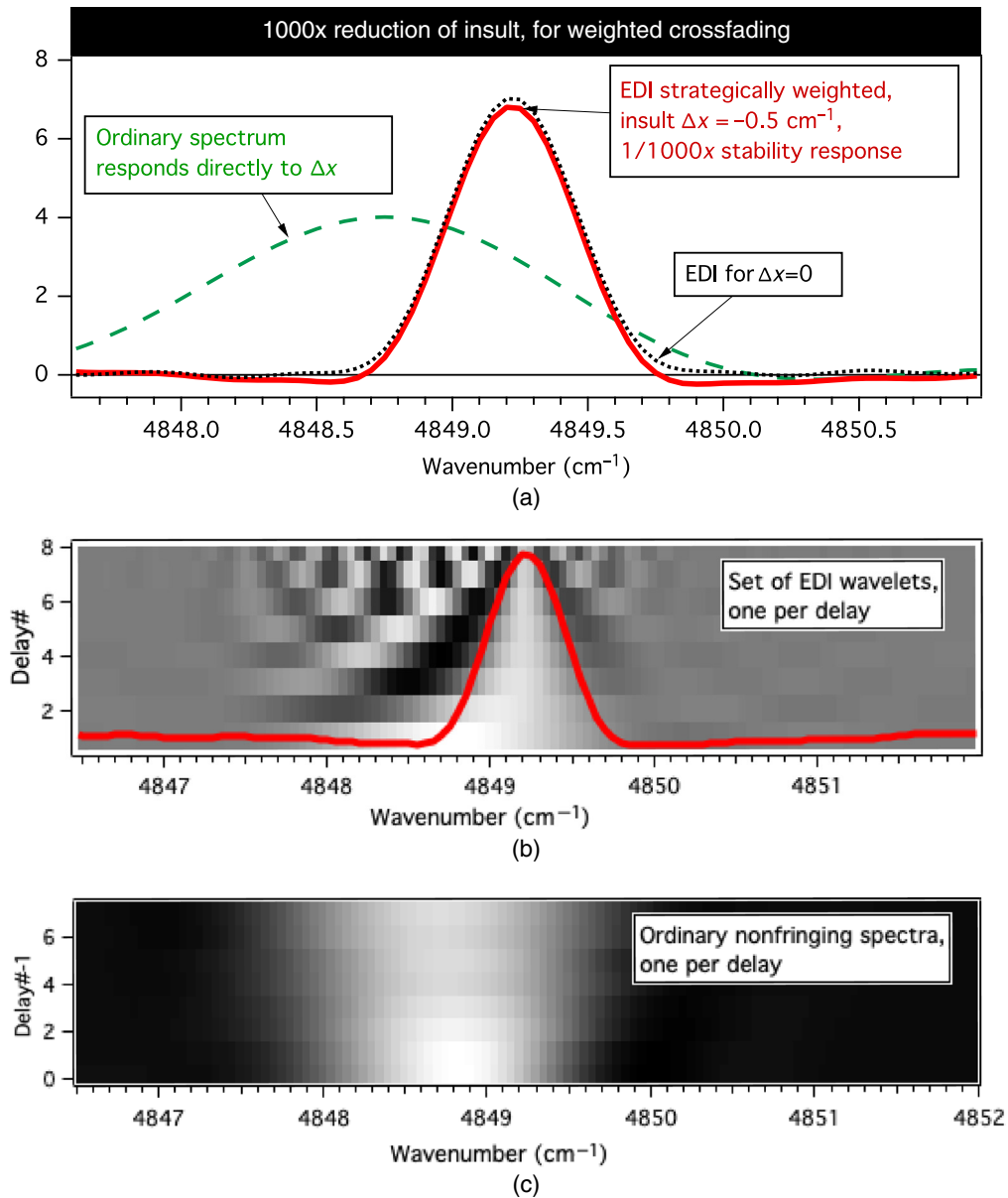


Fig. 26 Improved demonstration in 2019^{19,23} on the same T-EDI ThAr data as Fig. 25 produces 1000x stability for the net EDI peak (red curve) under an artificial insult Δx . This is 50x better than the 20x result of 2016 (Fig. 25) due to weights chosen to produce vector cancellation. The weights were calculated from neighboring 4764 cm^{-1} line acting as ersatz calibrant, then applied to 4849 cm^{-1} subject line here. About 0.001 cm^{-1} peak shift was observed comparing $\Delta x = -0.5$ to $+0.5$ cm^{-1} , for $\text{TRC} = 0.001$. (a) Black dotted curve is $\Delta x = 0$ case. Ordinary spectrum (green dashes) responds directly to Δx . (b) Net EDI is a sum of wavelets, each having different delay (etalon). Wavelets are shown prior to equalization (which diminishes E7 and E8). (c) Ordinary spectrum from each delay channel. These shift directly with Δx .

The smaller number reduces the number of independent weights that must be selected by 140x to only about 18 values. (We use a delay range from 0 to 1.8 cm, and the FT of the 200 points yields a frequency spacing of 0.1 cm.) Processing the whole band would then involve scanning this smaller region of interest repetitively across the band.

A consequence of the small number of points is that the FT of the data is fairly clean and simply shaped. In contrast, the FT of the full band of the order is fuzzy or grass-like, due to the multitude of features interacting with each other.

8.7.2 Work in postheterodyned space not detector space

Analysis is now done in output wavenumber space rather than input wavenumber space, that is, working after, not before the heterodyning reversal has been applied to the moiré data. That way, we do not have to explicitly model any idiosyncrasies in the heterodyning process.

8.7.3 Use working code for Fourier reaction

Figure 27 shows a computational intermediate step in computing the optimal weights, which is to use our EDI SR working-code to compute the reaction to a Δx of 0.5 cm^{-1} on a nearby ThAr at 4764 cm^{-1} as an ersatz calibrant, rather than use an approximate model based on the small signal approximation. Using the working-code, the exact Fourier rotation is found even for large insults Δx and also naturally takes into consideration any unknown idiosyncrasies to the algorithm.

For each etalon, the working-code produces the reconstructed output spectrum contribution, and we computed the change in FT. The imaginary part of this change is plotted (same as angular change times the magnitude)—this is what needs to be cancelled between the pair of delays by the weight choice.

The S-like shape that passes through zero at the center is essentially the lineshape times a line ($\rho - \tau_n$) that passes through the center of each delay peak. (However, the lowest delays $E0$ and $E1$ at 0.083 and 0.34 cm have negative frequency energy reflecting into positive branch, which distorts this S-shape.) The highest two delays $E7$ and $E8$ are not used for crossfading (they get zeroed during the equalization process) since they do not overlap other delays. (These high etalons were used for the highest sensitivity Doppler measurements on M-stars during the T-EDI project.⁹)

8.7.4 Weights are found independently of neighbors

The weights are adjusted to achieve a balanced condition. This is done semimanually in the code for a given frequency ρ and pair of delays (screenshot in Fig. 28) by adjusting the weights while observing computed vector results. The screen shows a vector display of complex values of lower delay peak 1 (ρ) (red), higher delay peak 2 (ρ) (blue), and their weighted sum (black diamonds), for three drift Δx values of -0.5 , 0 , and $+0.5 \text{ cm}^{-1}$, each having its own labeled marker. The sum of red and blue vectors produce the black vector. The goal is to place the -0.5 , 0 , and $+0.5$ black sum vector points along the same angle to the origin. Parameters “EtnumLeft” and “EtnumRight” are software indices that select the delay pair in this case ($E5$ and $E6$).

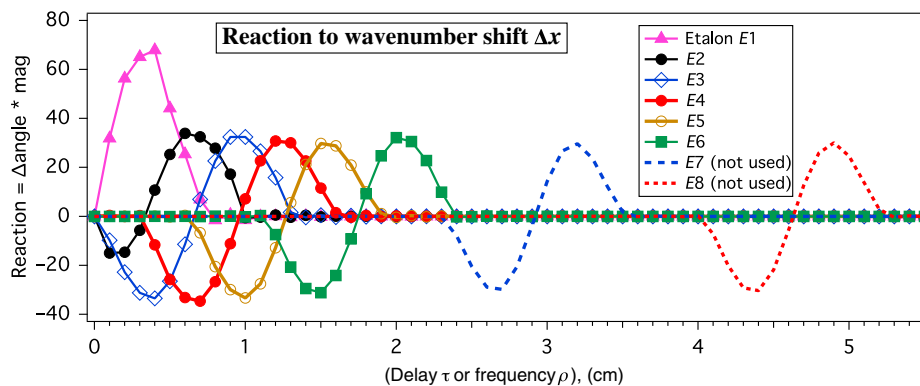


Fig. 27 Calculated reaction of the ersatz calibration line (4764 cm^{-1}) to a test insult $\Delta x = 0.5 \text{ cm}^{-1}$, for all the etalons of the T-EDI delay set. This is a computational intermediate step in computing the optimal weights, which are then applied to the subject line (4849 cm^{-1}). We used our working-code to produce the reconstructed output spectrum, then computed the change in FT. Their S-like shape, which passes through zero at the center, is essentially the lineshape times a line ($\rho - \tau_n$) that passes through the center of each delay peak, (except for the lowest delays $E1$, $E2$, where negative frequency energy reflects into positive branch). The vertical axis is reaction, defined as magnitude times angular change in complex plane. The highest two delays $E7$ and $E8$ are not used for crossfading since they do not overlap other delay.

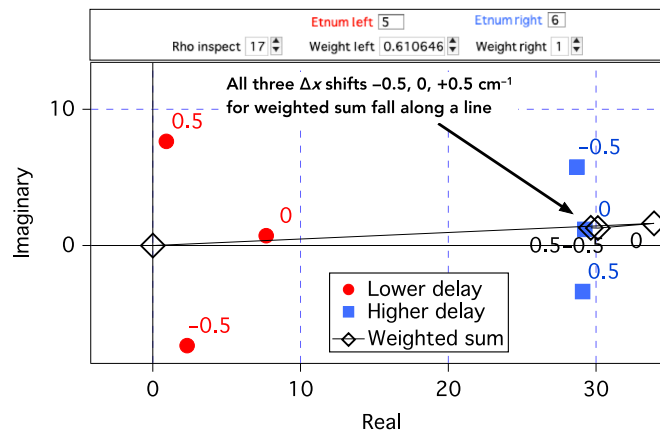


Fig. 28 Computer screenshot where we fine-tune selection of weightings to confirm phase cancellation, for a given frequency ρ and pair of delays. Vector display of complex values of lower delay peak₁(ρ) (red), higher delay peak₂(ρ) (blue), and their weighted sum (black diamonds), for three drift Δx values of -0.5 , 0 , and $+0.5 \text{ cm}^{-1}$, each having its own labeled marker. The goal is to place the -0.5 , 0 , and $+0.5$ sum vector points along the same angle to the origin. The value of ρ being examined is $17/10 = 1.7 \text{ cm}$ [parameter “Rho Inspect”]. Parameters “EtnumLeft” and “EtnumRight” are esoteric names of software indices that select the delay pair, in this case ($E5$ and $E6$). This solution for $\rho = 1.7 \text{ cm}$ corresponds to points FF in weight plot Fig. 29(b). It does not matter that the sum vector angle is slightly non-zero since that gets corrected during equalization.

Note that high and low peaks (red and blue) rotate with opposite polarity versus Δx . This is required behavior that allows one to steer the sum vector angle. The value of ρ being examined is $17/10 = 1.7 \text{ cm}$ [parameter “Rho Inspect”]. This solution for $\rho = 1.7 \text{ cm}$ corresponds to points FF in weight plot Fig. 29(b). It does not matter that the sum vector angle is slightly non-zero since that involves the combined signal postcrossfading and gets corrected during equalization.

Earlier, we sought lineshapes (such as triangle or sinc function) that would satisfy the small-angle crossfading Eq. (6). The concept of a shape implies that the result at one frequency is not independent of another neighboring one. Now in 2019, we evaluate the best weight ratio (w_1/w_2) of a low-delay high-delay pair (low and high) independently at each frequency ρ in between the delays. This is considered independently of the ratio at other frequencies—the overall or shape or symmetry of the line is not considered. We need not pay attention to the absolute size of w_1 or w_2 since this will be adjusted anyways during the subsequent equalization step.

Figure 29 shows the (a) unweighted lineshapes, (b) the weight solutions, and (c) the weighted lineshapes for T-EDI that produced our simulation shown in Figs. 1 and 26. The unweighted (a) and weighted (c) lineshapes are for information only and not used in the computation. Instead the weights (b) multiply the wavelets, which are then summed. These weights (b) are found in Fig. 28 to produce phase-shift cancellation by forcing the sum vector of each pair of overlapping reaction peaks (of a set $E1$ to $E6$) evaluated at each ρ , to maintain its original phase. (In principle, three or more delay peaks instead of two could be combined for crossfading. For simplicity, we consider here only two.)

This weight solution Fig. 29(b) is not unique. For example, the weight of the lower frequency peak was held (for convenience) at unity while varying the upper frequency weight. We are allowed to do this because only the ratio of weights matters at this point, and a subsequent equalization process will effectively further modify both weights to minimize ringing in the final PSF.

Note that weights for $\rho > 1.75 \text{ cm}$ (the $E6$ center) are forced to zero to prevent unbalanced reactions since there are no higher overlapping delays to cancel the reaction, and $E7$ and $E8$ are isolated not overlapped. The weight solution labeled FF at $\rho = 1.7 \text{ cm}$ is also shown as the vector diagram case in Fig. 28. Figure 29(c) weighted lineshapes are for information only, to show the namesake “X”-like crossings in between delays.

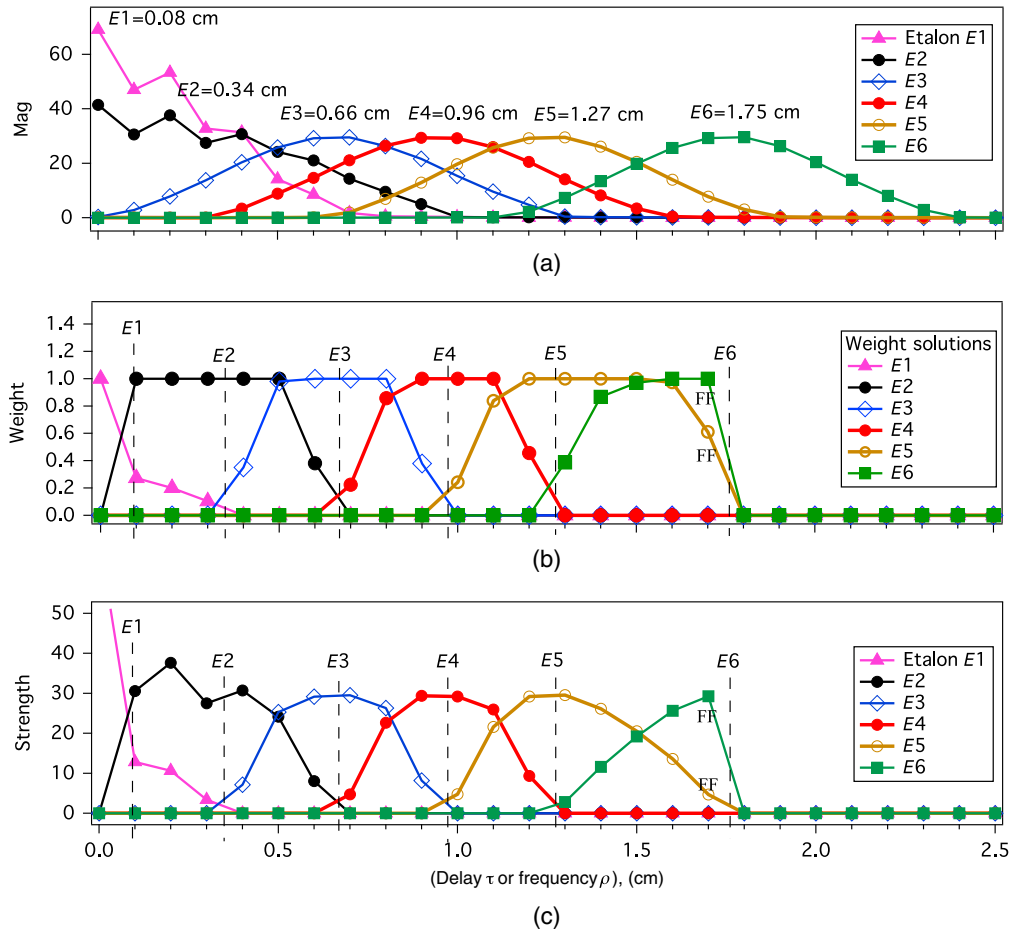


Fig. 29 Weight solutions of the T-EDI crossfading simulation: (a) unweighted lineshapes for information and (b) chosen weight solutions for the -0.5 to 0.5 cm^{-1} case of Δx . These produced phase shift cancellation by forcing the sum vector of each pair of overlapping reaction peaks ($E1$ to $E6$) evaluated at each ρ to maintain its original phase. Note that weights for $\rho > 1.75$ cm (the $E6$ center) are forced to zero to prevent unbalanced reaction since there are no higher overlapping delays to cancel the reaction, and $E7$ and $E8$ are not overlapped. Vertical dashed lines are labeled delay values. Weight solution labeled FF at $\rho = 1.7$ cm shown as vector diagram Fig. 28. (c) Weighted lineshapes for information only, to show the namesake “X”-like crossings in between delays. (c) The weighted and (a) unweighted lineshapes are not used in the computation—the weights (b) instead multiply the wavelets, which then get summed and equalized.

8.8 Equalization, Ringing, and Goal Resolution

For each delay, say $E1$, the moiré data are upshifted into a wavelet, the FT is taken of the wavelet, multiplied against the weights in Fig. 29(b) for $E1$, and then inverse Fourier transformed back to wavenumber space. This is repeated for the next delay, $E2$, and so on. Then all the wavelets are summed.

Then an equalization step is performed to remove bumps in the net MTF, based on how the 4764 cm^{-1} calibration line behaves during the same whole process, relative to an ideal Gaussian of width controlled by R_{goal} (here 9000). The user can choose R_{goal} to suite the needs of the application and input spectrum, with greater R creating greater ringing. (For example, Doppler velocimetry via fringe phase shifts would be less sensitive to ringing and would welcome the higher frequency information that would come from increasing R_{goal} . Although high-resolution spectroscopy to discern a small satellite line next to a much stronger line would desire to suppress ringing, lest it obscures the weak line.)

8.9 Results from Improved Crossfading Algorithm

Figure 30 shows the reaction $\delta\nu$ versus applied insult Δx for a range of insults, and for two different weightings schemes S and L, optimized for small (S) and large (L) insults of ± 0.2 and $\pm 0.5 \text{ cm}^{-1}$, respectively. (The L-weights are used in Fig. 29 and other figures.) The Y axis plots the change $\delta\nu$ in peak position for the sum of wavelets. Lines have slope of $\text{TRC} = \delta\nu/\Delta x$, which are reciprocal to stability gains $G_{\text{edi}} = \Delta x/\delta\nu$. The results confirm linear behavior $\text{TRC} \sim 0.001$ (dotted green line “1000x stability”) for small Δx . This is dramatically better than the 2016 results of 20x (green dot dashed line), and much better than a conventional spectrograph that defines 1x (black dashes), which is almost vertical it is so steep at this scale.

The behavior becomes non-linear at larger Δx approaching the half-width at half-max (HWHM) of EDI (open diamonds) or conventional native spectrograph (solid diamonds), 0.725 cm^{-1} and 0.27 cm^{-1} ($R = 3300$ and 9000). In these cases, we envision iterating using weighting L to first reduce the initial Δx to a smaller value before reapplying with weighting S. We evaluate the FTs numerically and do not yet have an algebraic expression for the peak location of the sum of wavelets. However, the second term of the Taylor expansion of the sine function is $\sin x = x - x^3/3! + \dots$, and we believe at large Δx the cubic term, which

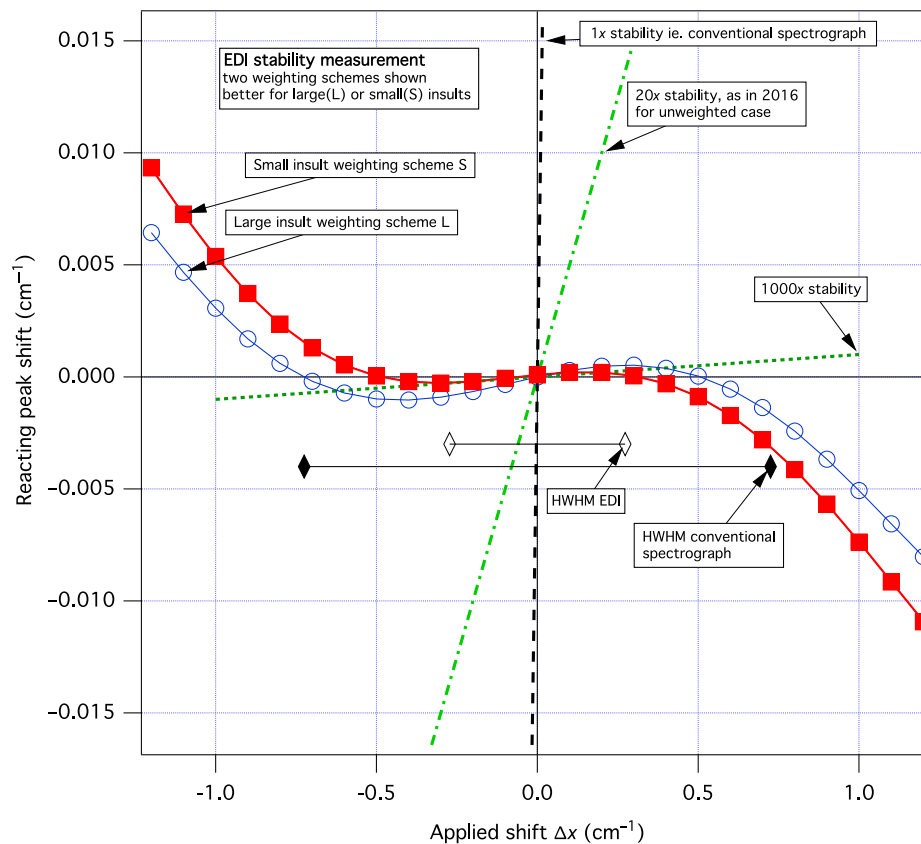


Fig. 30 Reaction $\delta\nu$ versus applied insult Δx for the crossfading simulation on T-EDI data, using two weighting schemes optimized for small (S, red squares) and large (L, blue open circles) insult ranges of ± 0.2 and $\pm 0.5 \text{ cm}^{-1}$, respectively. (The latter weights shown in Fig. 29.) Lines have slope of $\text{TRC} = \delta\nu/\Delta x$, which are reciprocal to stability gains $G_{\text{edi}} = \Delta x/\delta\nu$. The results confirm linear behavior $\text{TRC} \sim 0.001$ (dotted green line “1000x stability”) for small Δx . This is dramatically better than the 2016 results of 20x (green dot dashed line), and much better than a conventional spectrograph, which defines 1x (black dashes)—which is almost vertical it is so steep at this scale. Behavior becomes non-linear at larger Δx approaching the HWHM of EDI (open diamonds) or conventional native spectrograph (solid diamonds), 0.725 and 0.27 cm^{-1} ($R = 3300$ and 9000). In these cases, we envision iterating using weighting L to first reduce the initial Δx to a smaller value before reapplying with weighting S.

is negative, begins to dominate the linear term and is the reason the peak location curve bends downward (upward) for large (negative) Δx .

We see that for weighting choice S, better stability results for small Δx at expense of worse behavior for large Δx . We envision iterating, first using weighting L to reduce the initial Δx to a smaller value before reapplying with weighting S. The slight asymmetry of the graph is likely due to our use of experimentally measured data that may not have been perfectly restored to their $\Delta x = 0$ positions prior to the applied Δx .

8.9.1 Doppler-EDI simulation using single-pair crossfading on E4 to E5

An EDI used for Doppler velocimetry rather than general spectroscopy can use a single pair of delays, isolated and much higher in frequency than the native sensitivity peak. (Larger delays produce larger phase shift for a given Doppler velocity, and one does not need to measure the absolute shape of the spectrum.)

To simulate this, we follow in Figs. 31 and 32 what happens between the single pair of delays E4 and E5 during the crossfading, using the weighted solution already shown in Figs. 1 and 26, with weights Fig. 29. We ignore all frequencies but the ones in between E4 and E5, between 1.0 and 1.2 cm (labeling these “E4high” and “E5low”).

Figure 32 demonstrates that the low- and high-frequency sides of overlapping delay peaks E4 and E5 create wavelets that shift in opposite directions in response to Δx (a)–(c) that changes from -0.5 to 0 to $+0.5$ cm^{-1} . Having opposing shifts is a necessary condition for the cancellation, which is the essence of the crossfading method. Since the E4high and E5low wavelets

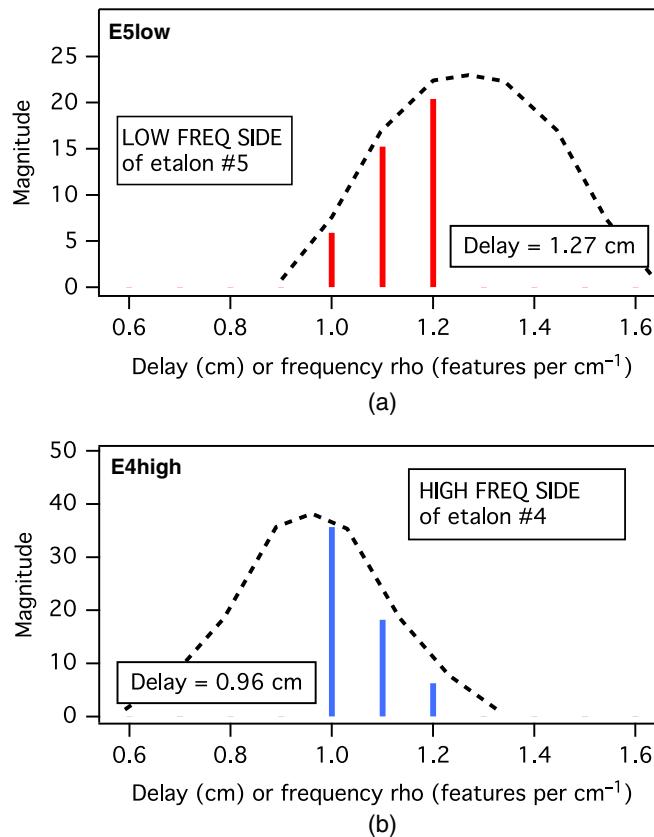


Fig. 31 We follow the results of a single crossfaded pair of T-EDI delays E4 and E5 isolated from the full delay set solution (Figs. 1 and 28) to simulate a Doppler EDI have a single-delay pair much higher than the native spectrograph peak. Here in Fourier space, we ignore all frequencies except those between 1.0 and 1.2 cm. The (a) low- and (b) high-frequency sides, called (a) E5low and (b) E4high, twist in opposite directions in complex space to an insult Δx . Although this is not apparent here (since we are plotting magnitude), it manifests in the splitting of wavelets (Fig. 32).

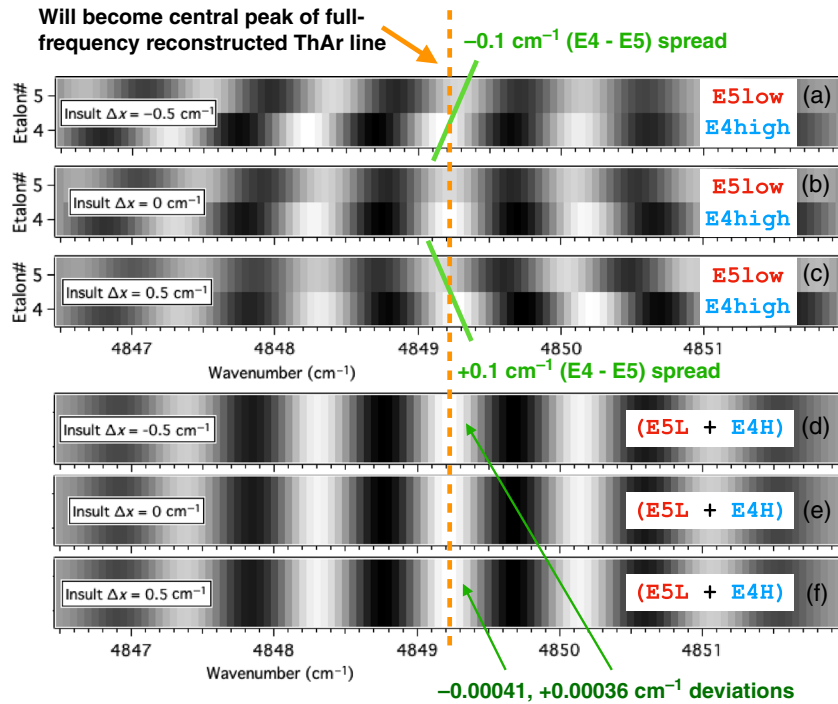


Fig. 32 Wavelets (ν -space) of crossfaded pair of delays $E4$ and $E5$ shown in frequency plot Fig. 31. Note that they shift in wavenumber space in opposite directions to Δx that changes (a)–(c) from -0.5 to 0 to $+0.5$ cm^{-1} , respectively. For $|\Delta x| = 0.5$ cm^{-1} , the splitting is 0.1 cm^{-1} between $E4_{\text{high}}$ and $E5_{\text{low}}$ wavelets at the central fringe 4849.2 cm^{-1} (orange dashes). This splitting or opposition is a necessary condition for crossfading to cancel a net reaction. (b) The neutral situation $\Delta x = 0$ shows $E4_{\text{high}}$ and $E5_{\text{low}}$ wavelets having matching phase at the central fringe. The central fringe becomes the central peak of the full-frequency (many-wavelet) reconstructed ThAr line. (The other, non-central fringes diminish away when the other wavelets are included because their many different phases do not align.) (d), (f) Summing the $E4_{\text{high}}$ and $E5_{\text{low}}$ wavelets to complete the pair produces vanishingly small net reaction (e). Central fringe positions are -0.00041 , 0 , $+0.00036$ cm^{-1} relative to the neutral case, for a TRC $\sim 0.5/0.0004 = 0.0008$ (1200x stability gain).

move in opposition, averaging them (d)–(f) can produce vanishingly small net reaction to Δx if the weights are chosen correctly. Positions of the central fringe are -0.00041 , 0 , $+0.00036$ cm^{-1} relative to the neutral $\Delta x = 0$ case, for a TRC of $0.5/0.0004 = 0.0008$ or 1200x stability gain.

9 Discussion

9.1 Iteration Useful for Large Insults

We envision that iterating one or two times would benefit crossfading when the insult is large. After the first pass, a large insult would be significantly reduced so that on a subsequent pass it would produce smaller phase angle shifts, which more accurately follow a linear behavior to the sine function (the imaginary part of $e^{i2\pi(\Delta x)(\rho-\tau)}$). This produces the best stability gain, working well over a range of Δx .

9.2 Use of Large Insult for Simulations

In our simulations here, we use a rather large insult of 0.5 cm^{-1} , which is similar to the full-width at half-maximum (FWHM) of the native spectrograph, in order to (a) make a shift graphically visible to the reader, (b) understand the boundaries of the performance parameter space at large Δx , and (c) confirm that the small Δx regime is linear (small angle approximation to the sine function).

9.2.1 For small insults, no iteration needed

For high-quality spectrographs already well mitigated with vacuum tanks, thermal control, and fiber-optic scrambling, the insult Δx would already be very small (much less than the native PSF FWHM), and thus in the linear regime where, theoretically, TRC can be zero. We anticipate a single application of crossfading will suffice in this linear regime, and the resulting stability gain due to the crossfading G_{edi} will further reduce the already small Δx by another 1 to 3 orders of magnitude. Exactly how close real software can get to theoretical performance has not yet been experimentally determined since we lack experimental data with sufficiently controlled Δx applied to all delay channels simultaneously.

9.2.2 For medium large insults, iteration suggested

For spectrographs not using conventional mitigations, the Δx could be medium large, similar to the native PSF FWHM. In an iterative fashion, the output spectrum could be compared to the input spectrum to measure the Δx that had occurred. With that knowledge of Δx , the input spectrum could be shifted conventionally to produce a first-corrected input spectrum. That new input spectrum, now having a greatly reduced Δx , could be reprocessed with crossfading but this time using a choice of weights optimized for smaller Δx (as in the solid red squares not open blue circles of Fig. 30) to produce a second-corrected output spectrum.

9.2.3 For largest insults, conventional translation

For the largest Δx greater than the native FWHM, the data would be translated in a conventional manner using calibration lines since then it is easy to accurately determine Δx . Then the resulting conventional repaired spectra would be processed with the crossfading algorithm.

9.3 Estimating the Small-Delay Performance

We estimate generally that crossfading can provide 10x to 1000x stability gain. The reason that the poor end of our estimate is only 10x is to include the configuration of a single delay of low delay value so that it overlaps the native spectrograph peak. (Example application would be to boost a spectrograph resolution $\sim 2\times$.) Since the EDI stability gain is roughly proportional to the delay value, a small delay has less “lever arm” for a given phase error due to mismatched weights or Δx .

We note that Fig. 27 shows the negative portion of “S-shaped” reaction to be reduced for the small delay $E2$ (0.34 cm) and absent for $E1$ (0.083 cm). We believe that this is related to using only a single-frequency axis for the FT display—for $E1$ and $E2$, the delay is low enough that the low-frequency side of their peaks lies on some portion of the negative frequency branch, and that there is reflection of signal energy from the negative frequency axis coherently adding to the positive frequency signals.

We will investigate this further in the future since the configuration of a single-delay EDI is of practical importance. Not only because it is simplest to build, but because if the EDI peak overlaps the native peak then crossfading can be performed for frequencies from zero to the center of the EDI peak, and importantly, the Δx is automatically the same for both native (lower frequency) and EDI (higher frequency) peaks for all time scales.

This work is preliminary. We currently lack an apparatus to make experimental measurements, similar to Fig. 17, where the same Δx can be precisely controlled or measured on a pair of delays. We are seeking funding to continue this exciting line of research.

9.4 Strategies for Time-Dependent Drift

The T-EDI era (2007 to 2011) data taking strategy of making many exposures for a given delay before moving to the next delay in the set was not ideal for crossfading. The long-time dwelling on the same delay, over and over again, unfortunately allowed Δx to wander under a time-dependent drift (about $0.4 \text{ cm}^{-1} \text{ h}^{-1}$ as shown in Fig. 36 of Ref. 11). We subsequently removed this bulk drift in the conventional way before using it in the crossfading simulation.

9.4.1 *Alternate between the delays*

If a simultaneous pair of delays (such as in scheme Fig. 17) is not available, and if instead we have a T-EDI style multiple-sequential delay interferometer in front of an echelle spectrograph, then we recommend alternating several times between the two delays of a pair, so that the effective time centroid is about the same for both delays.

Specifically, we suggest the following schedule kernel for two delays E_a and E_b , where D is the duration of a single exposure (at constant flux). Let the exposure time be D for E_a , then $2D$ for E_b , and finally return to E_a for another D . Then by symmetry arguments the time centroid of the combined two E_a delay data, and the E_b delay is both the same at $2D$ later than the beginning of the first exposure. This makes it most likely that any time-dependent delay will produce the same Δx for E_a and E_b net exposures. (We thank a reviewer for pointing out that there may be more uniformity in readout related noises and thermal issues to use two E_b delays in a row instead of a single E_b delay of twice duration.) Then one moves up the list pair by pair, measuring each pair with the above $D - 2D - D$ exposure kernel pattern. That way we are not comparing Δx at the very end of the highest delay with its value at the very beginning of the lowest delay. Instead, Δx would be judged (by crossfading) by what occurs within each pair. And since the time centroids are ideally overlaying, it should be near zero if it has linear time drift behavior.

9.4.2 *Drift has blurring effect on phase stepping*

To separate the fringing from non-fringing components of the EDI raw data, a minimum of three phase steps at roughly 120 deg or four at roughly 90 deg is needed. The effect of a linear time-dependent drift rate Rate on the total spread in time δt of the set of phase stepped exposures causes a spread by $\delta \nu$ due to the $\text{Rate} \times \delta t$. We believe that this acts like a rectangular blur of width $\delta \nu$ on the resulting fringing output, similar to opening up the spectrograph slit. In other words, a slow degradation of the fringe visibility.

For the T-EDI data, the phase stepping exposures were only a few minutes, and since we removed the slow bulk drift in a conventional manner (aligning spectral lamp calibration lines of all the exposures) prior to processing for the phase stepping, there was not significant blur over a phase step exposure due to drift rate. However if we had tried to make, say, hour long exposures per phase step, the drift during that time period could have been significant compared to the regular native blur—making the net blur larger and thus the effective lineshape in delay space narrower. This could change the choice of optimal crossfading weights.

9.4.3 *Discussion of interferometer stability and absolute spectral reference*

The interferometer comb can act like an intermediary calibrant, with a mathematically simple structure having only 3 degrees of freedom (apart from a small and known periodicity change across the band due to refractive index of glass). By the term “intermediary,” we acknowledge that the detailed value of the delay will drift slightly. But that is okay since the detailed value is determined by measuring the phases to an absolute calibrant such as an iodine cell or spectral lamp, especially if the calibrant is measured simultaneously with the star.

Using an absolute spectral reference, it is only necessary to casually stabilize the interferometer delay to about a quarter wavelength or better, to avoid smearing the fringes during a long exposure. Not to, say, $0.03/15000 \sim 2 \times 10^{-6}$ of a fringe, which is the 3-cm/s desired precision level of the Doppler measurement for a 1.1-cm delay at $0.54 \mu\text{m}$ wavelength (velocity per fringe proportionality of $c\lambda/\tau \sim 15 \text{ km/s per fringe}$). (Even if the delay is naturally very stable, such as with a Sagnac design,²⁴ we recommend using an absolute spectral reference.)

The two axes of possible interferometer cavity mirror tilt drifting over long exposures is less of a concern than the delay and easily remedied. One can use a TV to monitor the fringe appearance of a visible wavelength (such as HeNe laser) also passing through the cavity. The mirror tilt is then adjusted until the fringe appearance returns to normal, which can be done by remote actuators.

We are not concerned by slow drifts in the interferometer mirror tilt during long exposures, which could cause a slight change in the phase stepping. Each spectrum is analyzed independently, and precise phase steps are an output from the analysis—we only supply it very approximate

values to begin its search for a solution. Due to the massively parallel nature of fringing spectra (all the channels affected by the same step in delay), a precise solution for this delay step results from the fitting, and these steps can be irregular.

9.5 Proper Doppler Behavior Still Obeyed

Although it is good to show something is zero, we also show it is non-zero when it needs to be. We confirmed that our software including crossfading still produces in the output the proper Doppler shifted peak when the moiré phase is simulated to change to a simulated Doppler shift. We simulated it by applying a moiré phase change in proportion to each delay as $2\pi\tau\Delta\nu_D$ in radians (in addition to an ordinary wavenumber shift of the data).

Note that it is the moiré phase shifts that cause the EDI result to shift not the horizontal position along the dispersion axis. The moiré phase shifts only occur when the input spectrum moves relative to the interferometer periodic comb. Thus an unwanted instrumental drift, which shifts both the comb and the input spectrum, does not create the moiré phase shift. Hence the stable behavior of the EDI.

10 Application of EDI for Integral Field Spectrographs

10.1 For Gemini Planet Imager Integral Field Spectrograph

Integral field spectrographs (IFSs) are ripe for enhancement by EDI because they typically are starved for pixels for spectroscopy since they allocate pixels for the 2D spatial dimensions. Figure 2(b) shows a portion of the IFS detector for the GPI, and how only a few pixels are available for each subspectra that is assigned to a location on the sky image via a lenslet array. The GPI IFS can suffer ~ 1 pixel drift due to gravitational vector changes.²²

An interferometer can be inserted into the beam prior to the IFS, and its phase stepping will affect each of the many subspectra simultaneously. Suppose the source spectrum within a lenslet prior to dispersion is uniform in character. Then for the dispersed subspectra, one can sum transversely across the 2D subspectra to create a 1D subspectra, in the same way that is done for the T-EDI multi-order data.¹¹ Each lenslet subspectra could be analyzed independently and return separate results for the phase step values. The algorithm outputs detailed phase step values that fit the data. (However, one may get more accuracy for determining phase steps by first pooling all the spectra. Once the phase step values are found, one can then reprocess each subspectra independently.)

If the source spectrum for a given lenslet varies spatially in two-dimensions prior to dispersion, then the data analysis becomes more complicated—but we suspect still tractable with effort. However, we have not investigated this line of research.

The fact that an IFS has few pixels does not by itself prevent high-resolution spectroscopy. In principle, using the uni-phase data taking mode, each pixel should be analyzable to produce a spectrum, independent of others. We can point to the FTS, which only has effectively one pixel, yet can produce very high-resolution spectra when it scans over a wide range of delays.

In a previous conference proceedings,¹³ we simulated the advantages of adding an EDI to GPI IFS for the purpose of detecting fine molecular lines in exoplanet atmospheres (should there be sufficient flux). We found that this can allow detection, in spite of the low ($R = 70$) resolution that prohibits detection with the IFS alone. Figure 33 shows that a useful strategy is not to fill the entire delay space with a multitude of delay values but to use just a handful of delays and position them in a group at a location where the FT of the molecular spectrum has an especially large signal. Using the Earth's telluric spectrum as a source model, we found positioning the delay group near 0.58 cm produced a relatively large moiré signal. Figure 34 shows that the magnitude of this signal was similar to that measured by an $R = 3850$ conventional spectrograph, which is 55 times larger than the native resolution of 70.

The several delay peaks in the group overlap and thus are a candidates for crossfading, which would mitigate the ~ 1 -pixel drift. This stabilization might be sufficient to allow Doppler measurements of the molecular lines based on the phase of the moiré. In other words, using the phase of the moiré not just its magnitude.

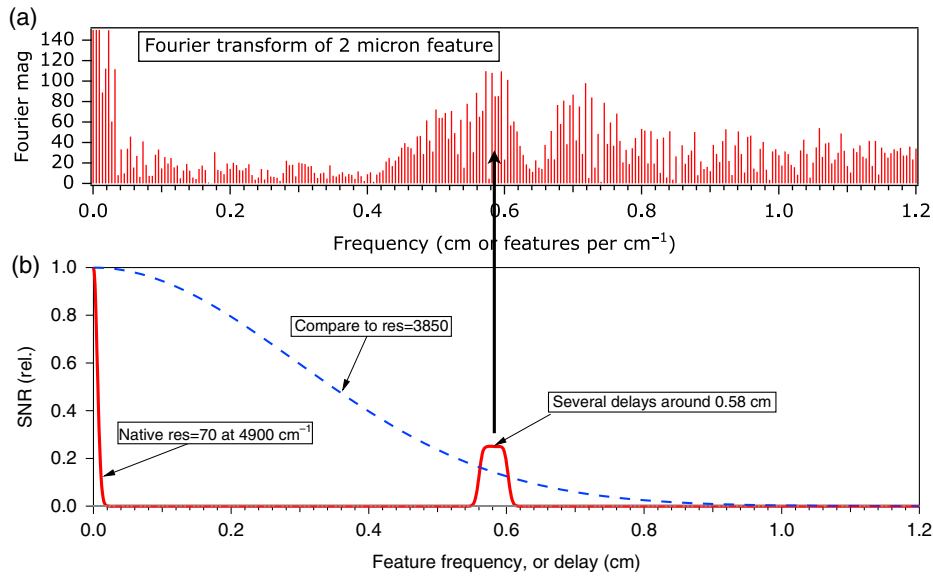


Fig. 33 (a) FT of a section of the CO₂ feature near 2 μm (4800 – 5100 cm⁻¹) showing concentration of energy above 0.4 cm and a local peak near 0.6 cm. (b) Red curve is filtering behavior, instrument response, or MTF of proposed EDI when a small set of delays is used to cover a region around 0.58 cm, and the native resolution is $R = 70$. Hence, the net peak is somewhat rectangular. A single delay would produce a peak of the same width as the native (at the origin) and of height 0.5. Using multiple delays spreads the area under the peak following a quadrature or root mean square sum rule where area under square of function is the same as a single delay, assuming the same exposure time is redistributed among several delays. The dashed blue curve is the response of a classical spectrograph of $R = 3850$, for comparison. Vertical axis plots signal relative to the same continuum photon noise, scaled such that both $R = 70$ and $R = 3850$ classical spectrograph cases have the same size pixels, and thus same noise in limit of broad features (zero feature frequency or delay). Horizontal axis is feature frequency or interferometer delay, which both have units of cm. Graphics reproduced with permission from Ref. 13 by SPIE.

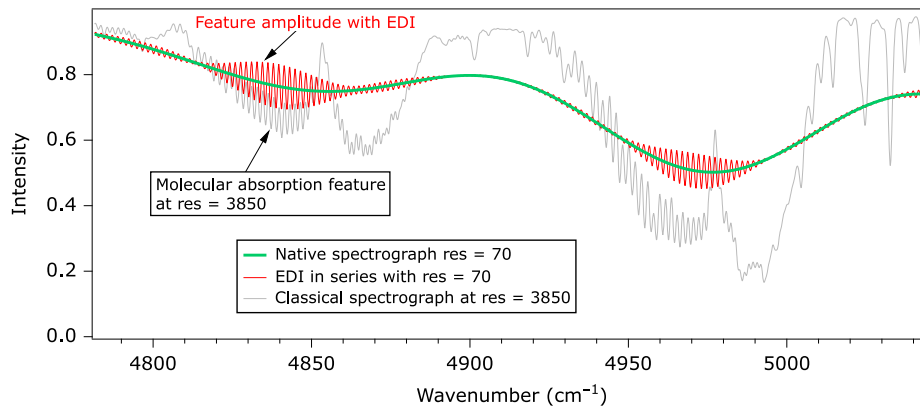


Fig. 34 Simulated measurement of the telluric spectrum using three calculated instrument responses to produce three curves: native $R = 70$ (green), classical $R = 3850$ (gray), and EDI (red). The semirectangular peak MTF of the bottom pane of Fig. 33 is multiplied against the FT of the telluric spectrum (top pane), then inverse Fourier transformed to wavenumber space. We use the localized native peak at 0 cm, the entire blue dashed curve, and the localized EDI peak at 0.58 cm to produce the green, gray, and red results here. The $R = 3850$ classical case (gray) resolves some of the fine lines enough for detection of this feature. However, the $R = 70$ native (green) cannot. The EDI peak at 0.58 cm produces the red curve, which produces a similar amplitude of fine lines of the CO₂ feature as the gray curve for some portions of the spectrum. This is noteworthy because the EDI boosts the low $R = 70$ resolution to effectively higher resolution to make a measurement otherwise not possible for this device. Graphics reproduced with permission from Ref. 13 by SPIE.

10.2 For Keck OSIRIS Integral Field Spectrograph

Figure 35 shows a simulation of how an EDI added to the Keck OSIRIS IFS would enable it to more sensitively measure the rotational broadening of an exoplanet (modeled at 25 km/s), by comparing the magnitude of the spectral FT at two delay values. These optimally are on the shoulder of the envelope ~ 1.5 cm in frequency (features per cm^{-1}). This measures the ratio of Fourier components between the exoplanet signal and a star (assumed to be rotating much more slowly at 3 km/s). The 1.5-cm frequency position has a large slope and is thus sensitive to changes in rotational velocity. The native OSIRIS having $R = 3800$ without the EDI cannot

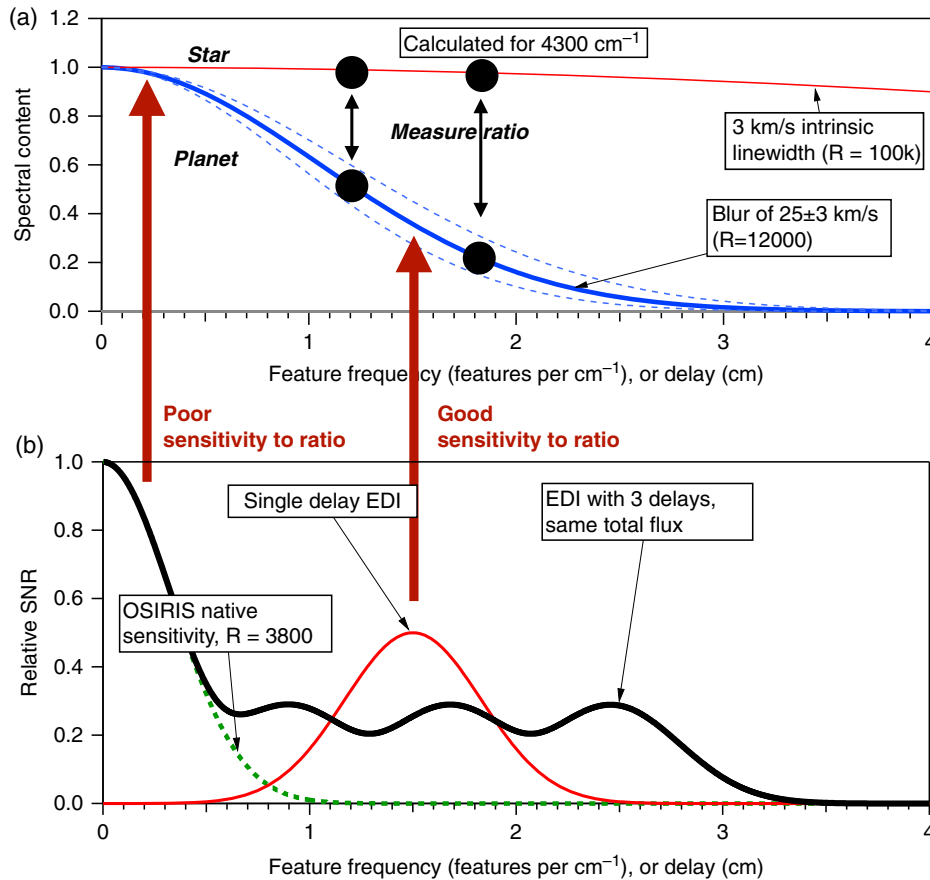


Fig. 35 (a) Doppler rotational broadening of an exoplanet can be measured by the decrease in fringe visibility ratio between star (red) and planet (blue), at delays (frequencies) near 1.5 cm. This is mathematically related to the autocorrelation technique used by Snellen.²⁰ In Fourier space, the effect of 25 km/s rotation at 4300 cm^{-1} is approximated as $R = 12,000$ Gaussian broadening. (b) OSIRIS $R = 3800$ native resolution (green dots) has sensitivity up to 0.7 cm, but this is too small to probe visibility ratio where it deviates from star above 1 cm. But an EDI with 1.5 cm delay makes a sensitive measurement (red peak) in the region where the blue visibility curve has a significant slope. Three delays could also be used (black curve), and the root-mean-square area under the EDI curves follows a constant area rule¹² for photon SNR for different delay spacing choices, where the area under the square of the red curve is redistributed to the fringing portion of the black curve squared [the portion above 0.7 cm not due to the native peak (green dots)]. Readout noise is not included in this graph and would lower the EDI performance. On the other hand, the EDI's robustness to FP noise could improve its relative performance (e.g., Fig. 36 on T-EDI data using NIR detectors having significant FP noise). These issues were intended for study in the (June 2019) Keck Instrument Call proposed feasibility study that produced this figure. We point out now that crossfading analysis could be applied to the three overlapping delays to dramatically increase their net stability in measuring the shape of the exoplanet spectrum, and thus in retrospect the three delay configuration is superior to the single delay (if wavenumber drift Δx dominates readout noise). This was not mentioned in the extremely brief white paper.

measure Fourier components above 0.7 cm. Hence the inclusion of an EDI can bring a new capability to OSIRIS at reasonably low cost.

11 Readout and Fixed Pattern Noise

Because the phase stepping (requiring at least three exposures per delay), there is more readout noise for EDI. The issue is analyzed in depth in Ref. 12, which considers photon noise, detector noise, and readout noise.

In many situations, instrumental noises such as drift (which EDI mitigates with crossfading) or fixed pattern (FP) noise (which EDI mitigates with phase stepping) dominate over readout noise. A real-world example of FP noise dominating readout and photon noise is Fig. 36 showing T-EDI measured spectra. The NIR detectors used by T-EDI are known to have significant FP noise compared to visible light detectors. In this example, the conventional spectrograph used alone (green curve) cannot detect the two ThAr lamp features (black dotted), but the EDI (red

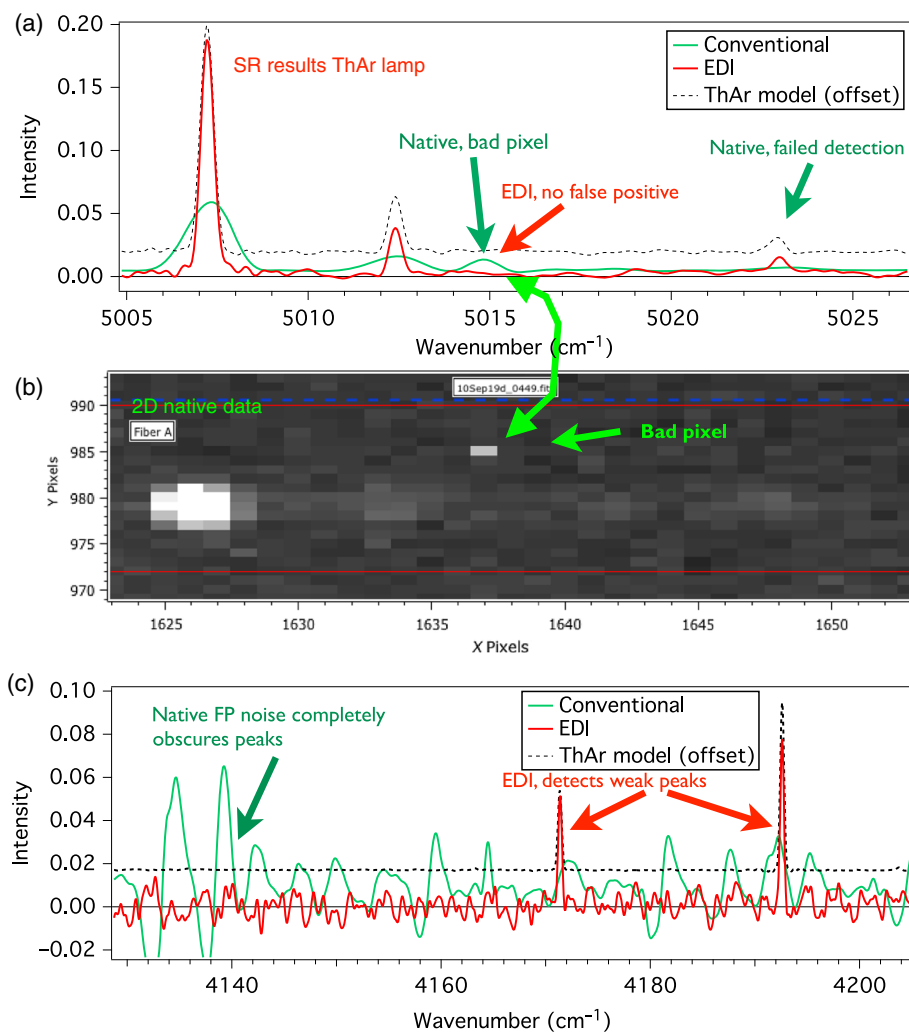


Fig. 36 Robustness to (c) FP noise and (b) bad pixels demonstrated in EDI's spectrally reconstructed output [red curves in (a) and (c)]. These instrumental effects can plague low light performance for conventional spectrographs. The near-infrared detectors used by T-EDI have significant FP noise. But because EDI phase stepping compares at least three exposures, any constant pattern is automatically cancelled. (c) T-EDI observes a ThAr lamp near the lowest ν of its range. Noise in the TripleSpec²⁹ native spectrograph (green curve) used alone obscures the two ThAr lines (peaks in black dotted theory curve³¹), whereas EDI (red curve) detects the lines easily. Graphics reproduced with permission from Ref. 11 by SPIE.

curve) can detect the two lines beautifully. The EDI taking data in a phase stepping manner act as a natural flat fielding process—only features that respond synchronously to the phase dither are detected—any constant background pattern is cancelled.

12 Conclusions

All spectrographs suffer drift insults of various kinds, and the crossfading-EDI technique reduces the reaction to insults in the output spectrum dramatically by moving the duty of fine wavelength determination from the dispersive spectrograph to the interferometer.

An interferometer can be added in series with most dispersive spectrographs to form an EDI. We recommend having at least two delays to perform crossfading during analysis. These improvements can boost the performance of the system dramatically over the spectrograph used alone. The stability is boosted from one to three orders of magnitude, and this boost multiplies stability gain G_{conv} from conventional mitigations (which reduces Δx), as $\delta\nu_{\text{out}} = \Delta x(1/G_{\text{conv}})(1/G_{\text{edi}})$. The spectral resolution can be boosted by a factor 2 to 20x or more, and the robustness to FP noise dramatically increased.

The crossfading EDI technique allows dramatically smaller, lighter spectrographs to achieve the same stability and resolution tolerances. This can benefit ultra-stable spectrographs for Doppler exoplanet characterization, long-term cosmic redshift measurements, improved stability for compact spectrographs used for Raman spectroscopy and biomedical analysis, and for remote sensing spectrographs on air- and space-borne platforms where weight and volume are critically limited.

The downsides are increased number of exposures (more readout noise), increased parasitic reflections from more optics (which diminishes flux and increases photon noise), and more complicated data analysis. However, in many conventional spectrographs, the precision is not limited by photon noise but dominated by PSF drift Δx . In this case, the small sacrifice in flux in using EDI reduces the net noise by removing a dominating drift noise component.

Results of this technique will greatly improve spectral precision and stability for at least two important, and complementary, methods of habitable exoplanet search currently limited by instrument stability or spectral resolution: (1) Doppler radial velocimetry using high-resolution spectrographs, by removing instrument noise so that the net precision is limited only by photon noise and (2) direct planet imaging using adaptive optics (such as the GPI²¹) that could analyze exoplanet atmospheres with a low-resolution integral field multi-object spectrograph, boosted in both resolution and stability by a crossfading-EDI.

Acknowledgments

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and based upon prior work supported by the National Science Foundation under Grant Nos. AST-0505366 and AST-096064, and NASA (Grant No. NNX09AB38G). An abbreviated version of this article was published in proceedings to the SPIE Conference on Astronomical Telescopes and Instrumentation 2020.³³ The authors would like to thank Ed Moses for his valuable support during the genesis years; Eric Linder, Chris Manning, and Tim Erskine for spurring me on; and Jerry Edelstein, Michael Feuerstein, Alex Kim, James Lloyd, Bruce Macintosh, Philip Muirhead, Matthew Muterspaugh, Richard Ozer, Martin Sirk, and Ed Wishnow for many fruitful conversations over the years. The author was inspired into the sciences by his father, John R. Erskine, recently passed, who showed me the beauty and wonder of the night sky.

References

1. D. A. Fischer et al., “State of the field: extreme precision radial velocities,” *Proc. Astron. Soc. Pac.* **128**, 066001 (2016).
2. D. J. Erskine, “Combined dispersive/interference spectroscopy for producing a vector spectrum,” U.S. Patent No. 6,351,307 (2002).

3. D. J. Erskine and J. Ge, “Novel interferometer spectrometer for sensitive stellar radial velocimetry,” in *Imaging the Universe in Three Dimensions: Astrophys. Advncd. Multi-Wavel. Imaging Devices*, ASP Conference Series, W. van Breugel and J. Bland-Hawthorn, Eds., Vol. 195, pp. 501–507 (2000).
4. D. J. Erskine, “An externally dispersed interferometer prototype for sensitive radial velocimetry: theory and demonstration on sunlight,” *Proc. Astron. Soc. Pac.* **115**, 255–269 (2003).
5. D. J. Erskine et al., “High-resolution broadband spectroscopy using an externally dispersed interferometer,” *Astrophys. J.* **592**, L103–L106 (2003).
6. D. J. Erskine and J. Edelstein, “Interferometric resolution boosting for spectrographs,” *Proc. SPIE* **5492**, 190–199 (2004).
7. J. Edelstein and D. J. Erskine, “High resolution absorption spectroscopy using externally dispersed interferometry,” *Proc. SPIE* **5898**, 297–307 (2005).
8. J. Edelstein et al., “Dispersed interferometry for infrared exoplanet velocimetry,” *Proc. SPIE* **7014**, 70147F (2008).
9. P. S. Muirhead et al., “Precise Stellar radial velocities of an M Dwarf with a Michelson interferometer and a medium-resolution near-infrared spectrograph,” *Proc. Astron. Soc. Pac.* **123**, 709 (2011).
10. D. J. Erskine et al., “Ten-fold spectral resolution boosting using TEDI at the Mt. Palomar NIR Triplespec spectrograph,” *Proc. SPIE* **8146**, 81460M (2011).
11. D. J. Erskine et al., “High-resolution broadband spectroscopy using externally dispersed interferometry at the Hale telescope: part 1, data analysis and results,” *J. Astron. Telesc. Instrum. Syst.* **2**, 025004 (2016).
12. D. J. Erskine et al., “High-resolution broadband spectroscopy using externally dispersed interferometry at the Hale telescope: part 2, photon noise theory,” *J. Astron. Telesc. Instrum. Syst.* **2**, 045001 (2016).
13. D. J. Erskine et al., “Enhanced exoplanet biosignature detection from an interferometer addition to low resolution spectrographs,” *Proc. SPIE* **10702**, 107024G (2018).
14. D. J. Erskine, “Dispersed interferometers,” in *The WSPC Handbook of Astronomical Instrumentation*, D. N. Burrows, Ed., Vol. 3, World Scientific, Singapore (2021).
15. J. Ge et al., “The first extrasolar planet discovered with a new-generation high-throughput Doppler instrument,” *Astrophys. J.* **648**, 683–695 (2006).
16. B. Ma et al., “Very low-mass stellar and substellar companions to solar-like stars from MARVELS. VI. A giant planet and a brown dwarf candidate in a close binary system HD 87646,” *Astron. J.* **152**, 112 (2016).
17. J. C. van Eyken, J. Ge, and S. Mahadevan, “Theory of dispersed fixed-delay interferometry for radial velocity exoplanet searches,” *Astrophys. J. Suppl.* **189**, 156–180 (2010).
18. A. G. Kim et al., “Giving cosmic redshift drift a whirl,” *Astropart. Phys.* **62**, 195–205 (2015).
19. D. Erskine et al., “Direct acceleration: cosmic and exoplanet synergies,” *Bull. Am. Astron. Soc.* **51**, 53 (2019).
20. I. A. G. Snellen et al., “Fast spin of the young extrasolar planet beta Pictoris b,” *Nature* **509**, 63–65 (2014).
21. B. A. Macintosh et al., “The Gemini Planet Imager: from science to design to construction,” *Proc. SPIE* **7015**, 701518 (2008).
22. S. G. Wolff et al., “Gemini Planet Imager observational calibration XIII: wavelength calibration improvements, stability, and nonlinearity,” *Proc. SPIE* **9908**, 990838 (2016).
23. D. J. Erskine and E. V. Linder, “A 1000× stabler spectrograph using an interferometer with crossfaded delays,” in *OSA Opt. Sens., Fourier Transf. Spectrosc. Top. Meeting*, Optical Society of America, p. FW5B.3 (2019).
24. R. Wei et al., “Design and experimental test of a common-path coherent-dispersion spectrometer for exoplanet searches,” *Publ. Astron. Soc. Pac.* **132**, 015003 (2019).
25. B. B. Behr et al., “Stellar astrophysics with a dispersed Fourier transform spectrograph. II. Orbits of double-lined spectroscopic binaries,” *Astron. J.* **142**, 6 (2011).
26. J. Harlander, R. Reynolds, and F. Roesler, “Spatial heterodyne spectroscopy for the exploration of diffuse interstellar emission lines at far-ultraviolet wavelengths,” *Astrophys. J.* **396**, 730 (1992).

27. J. Ge, D. J. Erskine, and M. Rushford, “An externally dispersed interferometer for sensitive doppler extrasolar planet searches,” *Publ. Astron. Soc. Pac.* **114**, 1016–1028 (2002).
28. R. Hilliard and G. Shepherd, “Wide-angle michelson for measuring doppler line widths,” *J. Opt. Soc. Am.* **56**, 362 (1966).
29. J. C. Wilson et al., “Mass producing an efficient NIR spectrograph,” *Proc. SPIE* **5492**, 1295–1305 (2004).
30. H. G. Roe, “Titan’s atmosphere at high resolution,” *Publ. Astron. Soc. Pac.* **115**(812), 1262 (2003).
31. F. Kerber, G. Nave, and C. J. Sansonetti, “The spectrum of Th-Ar hollow cathode lamps in the 691-5804 nm region: establishing wavelength standards for the calibration of infrared spectrographs,” *Astrophys. J. Suppl.* **178**, 374–381 (2008).
32. E. V. Linder, “Lineshape discussion,” Personal Communication (2014).
33. D. J. Erskine, “Boosting dispersive spectrograph stability 1000× using an interferometer with crossfaded delays,” *Proc. SPIE* **11451**, 114512D (2020).

David J. Erskine has been an experimental physicist at Lawrence Livermore National Laboratory since 1987. He has experience in (chronologically) femtosecond lasers, semiconductor physics, superconductivity, diamond anvil cell high-pressure physics, shock physics, high-speed recording techniques, Doppler interferometry, white light interferometry, Fourier signal processing, phase stepping algorithms, and digital holography. Since 1998, he has collaborated with astronomers to innovate interferometric techniques for the Doppler planet search and high-resolution spectroscopy. He is a member of SPIE.