Compression of intensity interferometry signals

Erez N. Ribak* and Yaron Shulamy

Department of Physics, Technion – Israel Institute of Technology, Haifa 32000, Israel

Abstract

Correlations between photon currents from separate light-collectors provide information on the shape of the source. When the light-collectors are well separated, for example in space, transmission of these currents to a central correlator is limited by band-width. We study the possibility of compression of the photon fluxes and find that traditional compression methods have a similar chance of achieving this goal compared to compressed sensing.

Keywords: interferometry; intensity interferometry; space interferometry; compressed sensing; photon counting.

1. Introduction

Intensity interferometry was born after the Second World War, following the development of radio technology. The method correlates the intensities from two separate light collectors pointing to the same object. It can be explained classically as sensitive to the deviations of the object's intensity from its mean, arriving at different times at the two collectors, depending on its geometry (Brown 1974). The quantum explanation to the effect was described by Purcell, Glauber and Mandel, and lead to the development of quantum optics.

Stellar intensity interferometry was developed into an experiment in Narrabri, Australia, and ran until 1974. Unfortunately it required long integration times and large collectors, as compared to amplitude interferometry, also developed at the same time. Both methods correlate signals from separate telescopes, but intensity interferometry performs second order correlations, compared to the more robust first order correlations in amplitude interferometry (see Fig. 1 and Appendix A). However, there are regimes where the latter is quite limited, and this is due to the atmosphere. First, the atmosphere is turbulent, and turbulence affects more the shorter wave lengths, limiting amplitude interferometry mostly to the infra-red. In contrast, intensity interferometry is oblivious to turbulence, and at the same time is more effective for hotter objects (where there are fewer photons per mode) (Brown 1974, Trippe *et al.* 2014). Here the atmosphere intervenes again, as it becomes opaque in the ultra-violet. Today experiments are being developed to increase the number and area of light-collecting dishes by employing

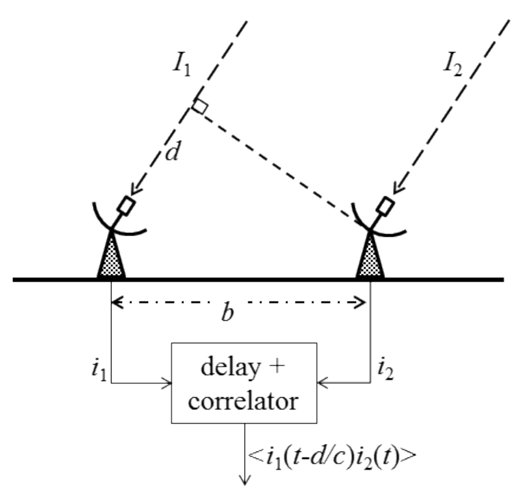

Fig. 1. Two dishes at baseline *b* collect stellar light. The photocurrents *i* represent the intensities, and are correlated, after correcting for the delay *d* between the beams. At low intensities the currents can be represented by single electron evens.

* eribak@physics.technion.ac.il

Cherenkov collectors (Dravins *et al*. 2012, Nunez *et al*. 2012). More such dishes can improve the correlation signal by using higher correlations (Ofir & Ribak 2006a, b).

It was proposed (Ofir & Ribak 2006b, Klein *et al*. 2007, Ribak *et al*. 2012) to move intensity interferometry observations outside the Earth's atmosphere, where requirements for amplitude interferometry are still beyond reach today. This can be realized by a formation flight: a fleet of satellites, each carrying a large, low-quality light collector. In space, objects can be tracked at any wave length for hours and days, necessary for intensity interferometry, and baselines (distances between satellites) can easily be varied, without real-estate constraints, and only at the accuracy of meters. These are only some of the reasons for the need to realize a complex system of this type.

When looking at the idea of intensity interferometry in space (and sometimes on ground) we see some new problems that should be taken into account. We chose to deal here with the limited ability of storing optical information digitally at each satellite, in order to transmit it when passing over a ground station. Another problem is transferring the information to a control center to perform the correlation between the intensities of the telescopes, in real-time or after the fact. Unlike amplitude interferometry, the signal from each station can be duplicated many times over, in order to correlate it with all other stations.

In amplitude interferometry (Labeyrie *at al.* 2006) the normalized correlation of two electric fields at two different locations and times is

$$\gamma^{(1)}\left(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2\right)=\frac{\left\langle E(\mathbf{r}_1,t_1)E(\mathbf{r}_2,t_2)\right\rangle}{\sqrt{\left\langle\left\|E(\mathbf{r}_1,t_1)\right\|^2\right\rangle\left\langle\left\|E(\mathbf{r}_2,t_2)\right\|^2\right\rangle}}\,. \quad (1)$$

This means that we need to have both fields hitting the same detector. The intensity correlation is similarly defined

$$\gamma^{(2)}\left(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2\right)=\frac{\left\langle I(\mathbf{r}_1,t_1)I(\mathbf{r}_2,t_2)\right\rangle}{\left\langle I(\mathbf{r}_1,t_1)\right\rangle\left\langle I(\mathbf{r}_2,t_2)\right\rangle}= \,. \quad (2)$$
$$=1+\left|\gamma^{(1)}\left(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2\right)\right|^2$$

Averaging is performed over integration time T, which is much longer than the wave-packet coherence time $\tau_{c.}$ If the baseline is $\mathbf{b} = \mathbf{r}_1 - \mathbf{r}_2$, and the difference in times is $\tau = t_1 - t_2$, we invoke ergodicity and define $\gamma\,(\mathbf{r}_1 - \mathbf{r}_2,\, t_1 - t_2) = \gamma\,(\mathbf{b},\, \tau) \equiv \gamma_b$. The intensity correlation is written in short as

$$\frac{\left\langle I_1(t)I_2(t-\tau)\right\rangle}{\left\langle I_1\right\rangle\left\langle I_2\right\rangle}=1+\frac{\tau_c}{T}\left|\gamma_b\right|^2 \,. \quad (3)$$

Here we can correlate the separately-measured intensities, simplifying the measurement and analysis.

During the early period of development of the intensity interferometry, all signal processing was mechanical or analogue (Brown 1974). In going to space, or even to remote earthly light collectors, data compression for storage and transmission is necessary, and we wanted to test its feasibility. We first examined the signal processing technique of compressed sensing, also known as sparse sampling (Elad 2010), which did not exist at earlier times. The motivation behind compressed sensing is that this method can be very useful in that the physics behind the signals is very well known, and if we look at the problem in terms of signal processing the signal can indeed be very sparse. If we know how to characterize signals in terms of physical behavior, we can build a dictionary consisting of several typical signals with known properties which can be used to represent the original signal. This is precisely the approach taken in the method of compressed sensing. To the best of our knowledge, temporal compression of interferometric signals was not proposed before, and it is our goal to see if indeed it can be useful.

Compressed sensing allows sampling a signal at a rate slower than the Nyquist limit. It relies on the signal being sparse in some basis Φ. Another condition beyond sparsity is that the signal is spread out in a conjugate space (e.g. frequency), and is very dense in Φ. A sensing method captures the information in the sparse signal and distills it into a small number of data points, even without knowledge of its content. Then optimisation is applied to reconstruct the full signal from its distilled representation. Both processes can be very fast.

Compressed sensing has long been used in astronomy, albeit not in the temporal sense. Radio-interferometric signals, coming from far-flung antennae or dishes, sample sparsely the Fourier plane of the celestial light distribution (as expressed in the Wiener-Khinchin theorem). Still, with knowledge of the physical constraints (e.g. object positivity and limited extent) it is possible to construct detailed maps of the object. Other examples abound.

Our goal in the following is to see if we can gain by compressing intensity interferometry signals, and whether compressed sensing is the preferred method.

2. Signal processing

In this study we performed a mathematical analysis of the signal characteristics. In a two-station intensity interferometer, the simplest case, we have two random signals, two photocurrents representing the two detected intensities, each having a Poisson distribution. Only after the correlation stage, done later and elsewhere, a new signal is obtained with known temporal properties, from which we can attain the stellar spatial coherence function that we pursue (Appendix A). So the mission is to collect the two electronic signals (electrical photocurrents), and compress them using the typical constraints: signal positivity (both stellar intensity and photocurrent are positive), and conservation of energy (the average photocurrent is constant and equal to the mean number of photons). The measurement noise we encounter, Poisson noise, obeys these constraints.

Once we understand the physical behavior of the signals, the most appropriate solution seems to be converting them from the time domain to the frequency domain: The later correlation of two signals in the time domain is equal to the product of the signals in frequency domain (Fig. 2). In this way, instead of sampling the temporal signal we sample the spectrum of the signal. This sampling in the Fourier temporal domain is similar to the just quoted case of radio interferometry in the object's Fourier spatial domain. Two such equally Fourier-sampled signals can be multiplied to obtain the Fourier-sampled transform of the correlation signal, at the same sampled frequencies. An inverse transformation, and application of the signal constraints, result in the temporal correlation. To fill in the gaps in frequencies we need some limitations on the correlation. We do not know much about the final result, except to say that it must be positive, have a narrow maximum somewhere near zero delay, and taper off to the mean flux at infinity. Unfortunately we cannot apply further symmetry arguments, as we do not know if the signal is even.

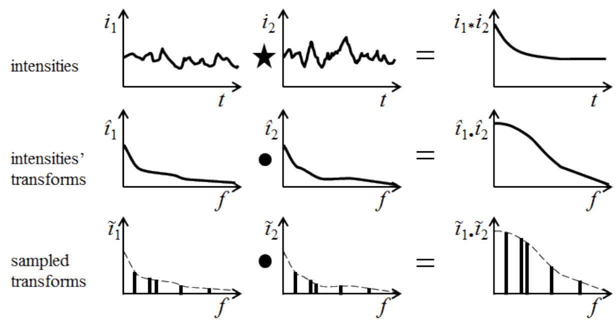

Fig. 2. Correlating two intensities (top row) is equivalent to multiplying their Fourier transforms (central row). In compressed sensing in the Fourier domain, when we sample the two signals at the same frequencies (bottom row), we get their product also sampled at the same frequencies. If we have enough prior (physical) knowledge about the object and noise, the gaps between the correlations samples might be bridged.

Thus we describe the correlation signal of mean-subtracted fluxes x_1, x_2 over a signal of length N as $y[n]=\frac{1}{N}\sum_{k=1}^{N}\left(x_1[m-n]-\overline{x}\right)\left(x_2[m]-\overline{x}\right)$, which is a δ function (representing the photons' correlations at times much longer than the coherence time, Appendix A). To that signal we add a Gaussian random process with no time correlation. In other words, we can write the correlation as

$$y[n]=\alpha\delta[n-n_0]+\beta z[n];$$
$$\alpha=\overline{x}^2\frac{\tau_c}{T}\gamma_b;\quad \beta=\frac{\overline{x}^2\tau_c/T+\overline{x}}{\sqrt{N}}. \tag{4}$$

Here n is the time bin, n_0 is the delay between signals, $z[n]$ is a random gaussian vector with zero mean and variance of 1, $\overline{x}$ the mean flux, τ_c the wave-packet coherence time, γ_b the coherence function, and T the integration time of one time bin, of which there are N such time bins (Appendix A). α is the sought signal and β is the accompanying noise term, that we want to reduce or filter out. So in terms of compressed sensing, the most appropriate dictionary representing the signal (a δ function) will be a dictionary of δ functions. From that dictionary we can fit a signal using the measurements, which in this case are supposed to be each a delta function that we want to measure. The advantage of the compressed sensing method is that we have to choose a limited number of blocks from the dictionary. In our specific problem the number of elements from the dictionary has to always be only one, which shows the fact that it is indeed a sparse signal. If our samples are taken in the frequency domain, the number of samples could be much lower than the number of samples we need in the temporal domain. The reason is that the location of the δ function (where the signals match in time) can be at any point on the temporal axis. However, if we take the samples in the frequency domain, the theory predicts that only one frequency sample is enough to find the function amplitude α, and more and more samples can make the solution less sensitive to noise. Notice that the original experiment (Brown 1974) used a basic lock-in detector at one frequency only to get rid of signal drift (different frequencies were used for the two beams).

We tried to figure out how much we can really dilute the number of samples in the frequency domain. We developed a model for our signal calculated its parameters to fit the theoretical expectations, and tested them by extensive simulations. The model (Appendix A) helps to understand the ability of compressed sensing to extract the signal from the noise and reduce the samples rate: we see (Fig. 3) that in the complex Fourier domain, at any frequency chosen, we get a circle of radius α, and addition of noise (growing noise term β, Eq. 4) hides this circle until such time that we cannot see its size any more.

3. Correlation of two random processes

Compressive sensing does not seem to fulfil the promise of efficient compression. This stems from the fact that the two photocurrents from the two light-collectors are Poisson processes, each without time correlation, and the true signal appears after their correlation. The information which we need for this correlation is distributed uniformly among all frequency samples simultaneously. When we take more samples in the Fourier domain, we can improve the signal to noise ratio. On the other hand, in the frequency domain the two signals are not correlated, and the intensity transforms are distributed uniformly. Our temporal process can be described as adding a δ function and a Gaussian process (Eq. 4 and Appendix A). In these two types of processes, the amplitude of each Fourier coefficient is independent of frequency. This means we have no room for manipulation in the frequency domain, because if the noise will be reduced so would the signal that we want to measure, and this is because their amplitude profile is the same. Thus, if we take any linear transformation of the signal, the signal to noise ratio will not change and we cannot separate the two.

4. Known time delay

Compressed sensing makes use of the difference between signal and noise projected into some space where that difference is maximized. Until now we tried to use compressed sensing in the Fourier domain, but can we do better if we have prior information in the temporal domain? Since we have a good guess of the delay between the two intensities, we can limit the time where we perform the correlation to a period centered on this delay, containing only N' samples compared to the full N samples, $N' \ll N$. This decreases noise term β (Eq. 4 and Appendix A) by a factor ~ $(N'/N)^{1/2}$, during the correlation stage. However, we still cannot reduce the Fourier bandwidth of the transmitted signals: by putting a limit on time, we add correlations between the Fourier components of the signal and noise, and reduce the number of independent samples. This would not be so bad, if we did not reduce the accuracy of the amplitude measurement at the same rate. Thus we cannot use the usual advantage of compressed sensing when the signal and noise have different spectral behavior: in our case, there is no region in the Fourier domain where the contribution to the signal (a δ function) is different from that of a white noise. Brown (1974) already showed that the final signal to noise ratio depends on the Fourier

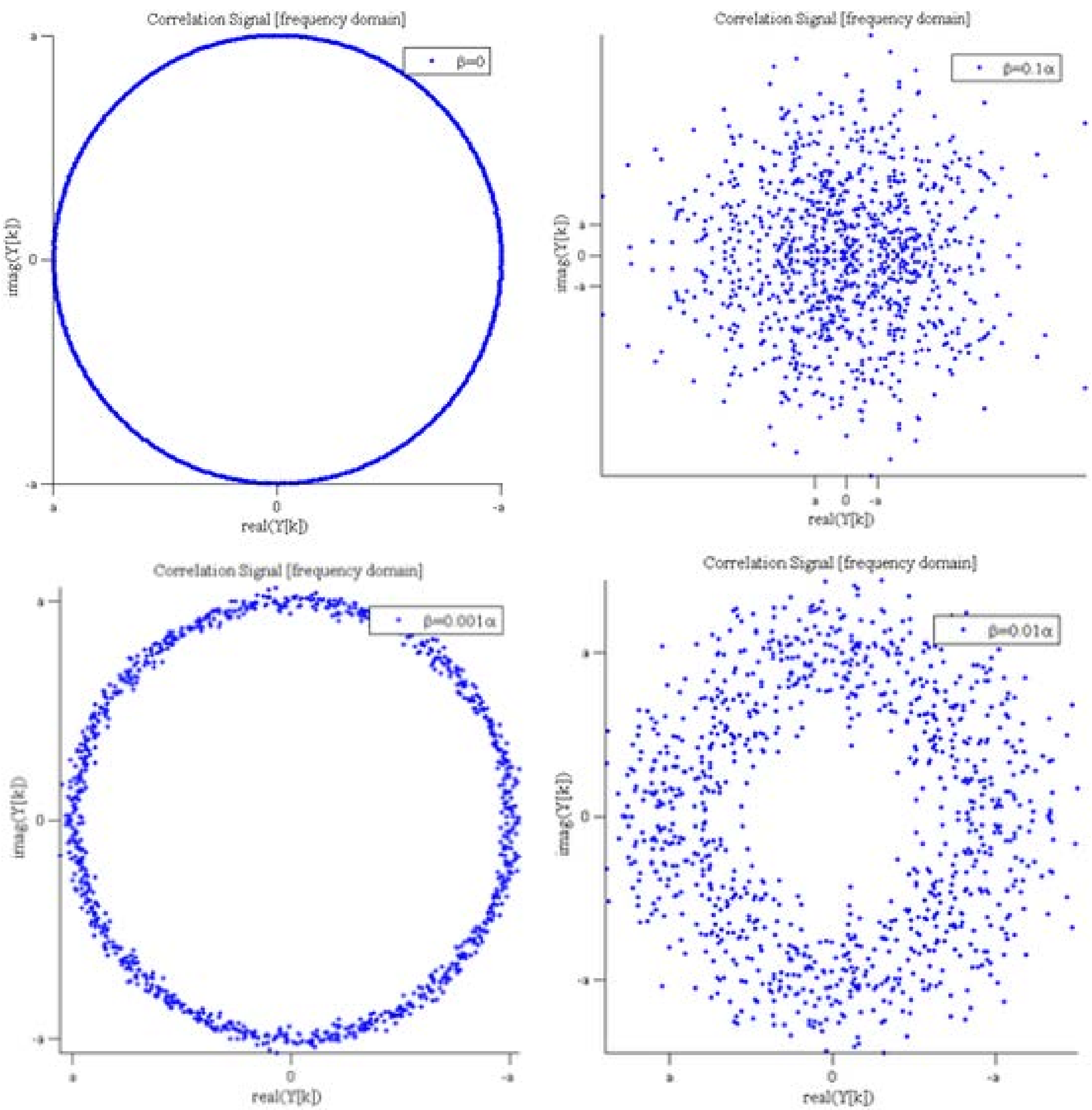

Fig. 3. The correlation signal in the complex (*Re. vs. Imag*.) frequency domain. Examples of sampling with noise values growing counterclockwise from top left. The values of β are each marked relative to the signal from 0 through 0.1 α, 0.01 α to 0.001 α. Simulation of N = 1000 measurements. The result is independent of the measurement frequency or of the temporal band.

bandwidth, without specifying from which frequencies is that bandwidth composed.

5. Low intensity signals

Until now, we looked at the problem of compressing the analogue signal (the photocurrent). For low luminosity objects we can represent the intensity signals, as sampled by the detectors, as sparse signals: modern photon counters produce a uniform pulse whenever a predefined noise threshold is crossed. The sparsity is represented here by a very low photon flux, namely long times between photon arrivals. This can also happen for brighter sources, if we separate the flux spectrally, with the advantage that each colour channel can be correlated separately with the corresponding colour channel at no loss (Brown 1974). While the average flux drops in each channel, the stellar size is assumed to be independent of colour and should add up with the other colours (and if not, we get information about their extended atmospheres at specific wave lengths). Where we can sample the signals sparsely, they are almost binary, and where they are not binary (i.e. having more than a single photon in a sample) it would be better to clip them and make them binary. This is because when we measure two electrons, there is much higher chance that they are due to noise than to two actual photons. Once the electric photocurrent is converted to a binary signal in the detector, we have to examine the possibility of using other compression methods. This is because compressed sensing restoration is complex and inaccurate for long signals.

6. Digital Compression

Searching for the optimal compression, there are algorithms intended for binary signals, but we do not know in advance if these 1s representing photons are correlated during the earlier stages of recording and transmission. In cases where the arrivals are totally random no compression method is useful. However, when the number of photons dwindles (as in weak objects or in narrow spectral bands), then the signals are mostly zero, which adds some order into them. We tested prefix code compression methods, where the amount of bits used is smaller. The advantage is realized when the detector sampling rate is too fast compared with its recording on hard disks, and using this compression we can reduce the bit rate needed, and make the recording of the digital signal possible. When the signal is transmitted to the central station, this compression method allows a lower transmission rate.

When the flux $\bar{n}$ is low, we expect statistically to get long strings of zeros, separated by few ones. Because neighboring cells are not correlated and we measure the arrival or non-arrival of a photon in a time slot, we are dealing with a geometric distribution $p(k, \bar{n}) = (1 - \bar{n})^k \bar{n}$. For convenience we denote $P = \bar{n}$ as the probability of one (photon) and $Q = 1 - \bar{n}$ as the complementary distribution. The average string will be $\langle k \rangle = P^{-1}$ long, but we need to consider all string lengths.

For a large number of photons, we might be able to tolerate some losses in compression, as long as these losses do not contribute more than the inherent noise. However, here we choose to deal with lossless compression, and try a method which compresses only the number of zeros, by inserting a unary number of ones and a final zero before its binary representation: if the number is b digits long we have b-1 ones and a single zero. The number k = 150 or binary 10010110 (8 digits long) will be represented by seven ones and one zero followed by k, thus: 1111111010010110. Hence the 150 zeros and a single one following them was reduced into 16 bits. In general, each number k will be represented by b bits such that $b = 2\lceil \log 2(k+1) \rceil$, where $\lceil x \rceil$ denotes the ceiling value. To see the compression capability of a string of numbers with a constant geometric distribution, we calculate the average length of a compressed string to be (Appendix B)

$$\langle b \rangle = \sum_{k=0}^{\infty} b(k)\, p(k) = \\ = 2\left[P - \log_2 P\right] \underset{k>>1}{\cong} 2\left[\frac{1}{k} + \log_2 k\right]. \quad (5)$$

For a sparse signal, the mean number of bits will behave as the logarithm of the uncompressed signal. The compression ratio of a flux $\bar{n}$ will be

$$\mathrm{comp}\left(\bar{n}\right)=\frac{\langle b\rangle}{\langle k\rangle}=\frac{2\left[\bar{n}-\log_2\bar{n}\right]}{1/\bar{n}}= \\ =2\left[\bar{n}^2-\bar{n}\log_2\bar{n}\right]. \qquad (6)$$

If the transmission or recording rate is S, the highest sampling rate of the detector is S / comp ($\bar{n}$) = S / comp ($2.5^{-m} P\, A\, \eta\, T\, \delta\omega$) (Eq. A1). Today's detectors and amplifiers reach $\delta\omega \approx 1$ GHz, and at that speed we gain by compression starting at magnitudes fainter than $m > 8$, and at $m = 11.5$ compression is tenfold (Fig. 4).

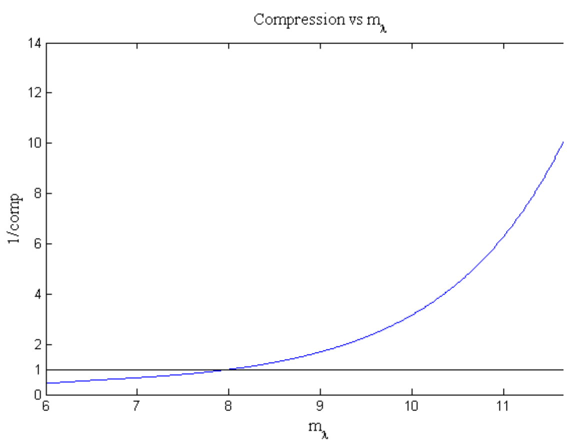

Fig. 4. The ratio of full signal to compressed signal as a function of stellar magnitude (or of brighter stars at narrow spectral bands). Compression is exponential and at 12th magnitude reaches a factor of 13, but it is not effective below 8th magnitude. Here

These results are calculated for the case of intensity interferometry, and show that it is possible to compress signals even from faint stellar objects observed in space (or distant ground locations). Spectroscopy allows employing many parallel channels without loss of signal, improving the sensitivity of intensity interferometry (Section 5). Relaying multiple channels to the central correlation station requires wide band, but the significant compression of each of the many weak channels will allow reducing that rate again.

In addition, compression can be performed for other astronomical photon-counting applications, such as speckle imaging, faint spectroscopy, wave front sensing and more.

7. Conclusions

In the case of stellar intensity interferometry signals, compression methods were found to be inefficient due to the special characteristics of the separate signals: without prior knowledge about which pieces of data are important for the later correlation, we cannot improve the sampling of the two signals. Since both signal and noise are spread evenly in time and in frequency, compressed sensing methods, which look for difference in their behavior, do not have any advantage. Essentially, all linear signal processing methods have hard time coping with the data, since the signal and noise information are statistically equally divided among the samples.

When we moved to low light levels we found a theoretical use for the compressed sensing method as a way to reduce the number of samples, but even in this case its seems that digital compression has more advantages over compressed sensing, and this is because when we count bits and not samples, plain digital compression can give much higher bit rate reduction in less time and without losing information on the way.

Acknowledgements are due to S. G. Lipson. The Israel Science Foundation and the Ministry of Science supported parts of this work.

References

Brown R H, 1974, *The Intensity Interferometer its Application to Astronomy*, Taylor & Francis, London.

Dravins D *et al*. 2013, Astroparticle Physics 43, 331.

Elad M, 2010, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York.

Klein I, Guelman M, & Lipson S G, 2007, Applied Optics 46, 4237.

Labeyrie A, Lipson S G & Nisenson P, 2006, *An Introduction to Optical Stellar Interferometry*, Cambridge University Press, Cambridge.

Nuñez P D *et al*. 2012, MNRAS 424, 1006.

Ofir A & Ribak E N, 2006a, MNRAS 368, 1646.

Ofir A & Ribak E N, 2006b, MNRAS 368, 1652.

Ribak E N, Gurfil P & Moreno C, 2012, SPIE 8445-8.

Trippe S *et al*. 2014, JKAS 47, 235.

Appendix A

We examine the case of a star of magnitude m_v at zenith, a collector area A, and detection efficiency η, for which the power is $P\ (m_v = 0) = 5{\cdot}10^{-5}\ \mathrm{sec}^{-1}\ \mathrm{m}^{-1}\mathrm{Hz}^{-1}$ (Labeyrie *et al*. 2006). Let $x_1[n]$ and $x_2[n]$ be the numbers of photons arriving at each detector during time n which is an integral over the interval T. Then the mean flux is

$$\overline{x} \equiv E\left(x_i[n]\right) = 2.5^{-m_v} P\eta A \frac{T}{\tau_c}, \tag{A1}$$

Since the two fluxes have the same mean and variance, their correlation variance is

$$\begin{aligned}\mathrm{var}\left(x_1[n]x_2[m]\right) &= E\left\{\left[x_1[n]x_2[m] - E\left(x_1[n]x_2[m]\right)\right]^2\right\} = E\left(x_1[n]^2 x_2[m]^2\right) \\ &= E\left(x_1[n]^2\right)E\left(x_2[m]^2\right) = \mathrm{var}\left(x_1[n]\right)^2\end{aligned} \tag{A2}$$

The covariance of the two photon fluxes is different from zero only for matching times, and the correlation expectation includes the sought coherence function,

$$\mathrm{cov}\left(x_1[n], x_2[m]\right) = \left(\frac{\tau_c}{T}\overline{x}^2 + \overline{x}\right)\delta\left[n-m-n_0\right] = \frac{\tau_c}{T}\overline{x}\gamma_b\delta\left[n-m-n_0\right], \tag{A3}$$

which does not average to zero only at time difference n_0. The correlation of the mean-subtracted fluxes over a signal of length N is

$$y[n] = \frac{1}{N}\sum_{k=1}^{N}\left(x_1[m-n]-\overline{x}\right)\left(x_2[m]-\overline{x}\right). \tag{A4}$$

Thus $y\ [n]$ can be thought of a mean of N random, zero-mean variables of the same variance. Since N is large ($10^{\sim 12}$), the mean of the sum will be the same, but the variance will be scaled down by $1/N$,

$$\mathrm{var}\left(y[n]\right) = \frac{\mathrm{var}\left(x_i[n]\right)^2}{N} = \frac{\left(\overline{x}^2\tau_c / T + \overline{x}\right)^2}{N}. \tag{A5}$$

For $n \neq n_0$ we get $E\ [y\ (n)] = 0$, but for $n = n_0$

$$E\left(y[n_0]\right) = \mathrm{cov}\left(x_1[n_0+n], x_2\left[n\right]\right) = \overline{x}^2\frac{\tau_c}{T}\gamma_b, \tag{A6}$$

which is a random gaussian function with known mean and variance. However, the position of n_0 is unknown, although it must be close to zero delay. In other words, we can write the correlation as

$$y[n] = \alpha\delta[n-n_0] + \beta z[n]; \quad \alpha = \overline{x}^2\frac{\tau_c}{T}\gamma_b; \quad \beta = \frac{\overline{x}^2\tau_c / T + \overline{x}}{\sqrt{N}}. \tag{A7}$$

Here $z\ [n]$ is a random gaussian vector with zero mean and variance of 1, the result of averaging $N \gg 1$ random processes. We have two degrees of freedom, the delay n_0 and the correlation signal α. Experimentally we try to overcome the noise by increasing the signal x and the number of measurements N.

The correlation of the two signals of length N (Eq. A3) can be described as the product of the two Fourier transforms of these signals,

$$Y[k] = \mathfrak{F}\left[\frac{1}{N}\sum_{m=1}^{N}\left(x_1[m-n]-\overline{x}\right)\left(x_2[m]-\overline{x}\right)\right] = \frac{1}{N}\left(\hat{x}_1[k]-\overline{x}\right)\left(\hat{x}_2[-k]-\overline{x}\right). \tag{A8}$$

If we wish to sample them in only a few frequencies, we must do that at the same frequencies, then multiply them and transform back to the time domain, to obtain their correlation (Fig. 2). We express the Fourier correlation as $Y = Y_s + Y_n$, namely the signal and noise components. The signal correlation will be

$$Y_s[k] = \mathfrak{F}\left\{\alpha\delta[n-n_0]\right\}_{[k]} = \alpha \exp\left(i2\pi k n_0 / N\right). \tag{A9}$$

We choose L Fourier terms, $\kappa \subseteq \{1, 2,,,N\}$, $|\kappa| = L \ll N$, the sampling vector will be $C[k]$ which is a subset of $Y_s[k]$. Here k is an index, mapping each of the group $\kappa[k]$ arbitrarily into the matching index,

$$C[k] = \sum_{n=1}^{N} y[n] \exp\left[i2\pi\kappa\left(k\right) n / N\right] \equiv Dy\left[n\right], \tag{A10}$$

where the elements of matrix D (also called the dictionary matrix) are given by $D_{kn} = \exp\,[i2\pi\kappa(k)n/N]$. We are seeking a solution giving a δ function, which means that it is of very low sparsity. We find it through minimization of $\|y\|_0$ subject to $Dy = C$, which is a typical compressed sensing problem with known solutions such as projection, matching pursuit, orthogonal matching pursuit, and other variations (Elad 2010). They all converge to the same solution when we only have to identify a single coefficient, by requiring the amplitude to provide the least error

$$\tilde{y}_n = \arg\min_{y_n} \left\| D_n y_n - C \right\|_2^2 . \tag{A11}$$

D_n is the nth column vector (of length L) in matrix D. The solution is

$$\tilde{y}_n = \frac{D_n{}^T C}{\left\| D_n \right\|^2}, \tag{A12}$$

whose maximum value is $L\,\alpha^2$, found at index

$$\tilde{n} = \min_n \left\| \left(C - D_n \tilde{y}_n \right) \right\|^2 = \min_n \left(\left\| C \right\|_2^2 - \frac{\left(D_n{}^T C \right)^2}{\left\| D_n \right\|_2^2} \right) = \max_n \frac{\left(D_n{}^T C \right)^2}{\left\| D_n \right\|_2^2} . \tag{A13}$$

We now turn to the noise, represented by $\beta\, z[n]$ (Eq. A7) with a Fourier transform $\beta\, Z[k]$. The variance at each frequency is

$$\operatorname{var}\left(Z[k]\right) = E\left(ZZ^*\right) = E\left(\sum_{n=1}^{N} \beta z[n] e^{\frac{i2\pi kn}{N}} \sum_{m=1}^{N} \beta z[m] e^{\frac{-i2\pi km}{N}} \right) = \beta^2 \sum_{n=1}^{N} E\left(z[n]^2 \right) = N\beta^2 , \tag{A14}$$

independent of frequency. Returning to the Fourier correlation, we have

$$E\left(Y[k]\right) = C[k] \le L\alpha^2; \quad \operatorname{var}\left(Y[k]\right) = N\beta^2 . \tag{A15}$$

The variance is a random variable made of a sum of N random complex gaussian variables. These variables can be written as

$$W[k] \equiv \beta \sum_{n=1}^{N} x[n] e^{i2\pi\kappa(k)n/N} = \beta \sqrt{\frac{N}{2}} \left(u + iv \right), \tag{A16}$$

where u and v are zero-mean real gaussian variables with unity variance. The combined signal $Y[k]$ is shown in Fig. 3.

Appendix B

The expectation of the number of bits of distribution k is

$$\langle b\rangle = \sum_{k=0}^{\infty} b(k)\, p(k), \tag{B1}$$

where for $k = 0$, the number of bits is 2, and for every other value it follows the ceiling (next integer) of the logarithm of k

$$b(k) = 2\left\lceil \log_2(k+1)\right\rceil. \tag{B2}$$

If $P = \bar{n}$ is the probability to have one photon during a time slot and Q is the probability to have no photon ($P + Q = 1$), then the expectation of each k is $p(k) = Q^k P$. Thus

$$\langle b\rangle = \sum_{k=0}^{\infty} b(k)\, p(k) = 2\bar{n} + \sum_{k=1}^{\infty} 2\left\lceil \log_2(k+1)\right\rceil Q^k P. \tag{B3}$$

Notice that the RHS summation starts at $k = 1$. If we define $k = 2^l - 1$, we get for that summation

$$\sum_{k=1}^{\infty} 2\left\lceil \log_2(k+1)\right\rceil Q^k P = \sum_{k=1}^{\infty} \lceil l \rceil Q^k P. \tag{B4}$$

Thus we can divide this sum into sub-sums where each is an addition from one power of two to the next. Thus the number of bits in each such sub-sum is constant and can be taken out of the total,

$$\langle b\rangle = 2P + 2P\sum_{l=0}^{\infty} l \sum_{k=2^l-1}^{2^{l+1}} Q^k. \tag{B5}$$

Each separate sum is an algebraic. Thus

$$\langle b\rangle = 2P + 2P\sum_{l=1}^{\infty} l \sum_{k=2^l-1}^{2^{l+1}} Q^k = 2P + 2P\sum_{l=1}^{\infty} l \frac{\left(Q^{2^{l+1}} - Q^{2^l-1}\right)}{P} = 2P + 2\sum_{l=1}^{\infty} l\left(Q^{2^{l+1}} - Q^{2^l-1}\right). \tag{B5}$$

As the number of consecutive zeros, k, grows at lower flux, the sum approaches very fast ($-\log_2 P$) and the approximate expectation of the number of bits will be

$$\langle b\rangle = 2\left[P - \log_2(P)\right] \cong 2\left[\frac{1}{k} + \log_2(k)\right]. \tag{B6}$$