NON-CONVENTIONAL DETECTOR APPLICATIONS FOR DIRECT FOCAL PLANE COVERAGE

M. Gai[1], M.G. Lattanzi[1], S. Casertano[2], M.D. Guarnieri[1,3]
[1]Astronomical Observatory of Torino, Str. Osservatorio, 20, I-10025 Pino Torinese, Italy
[2]ESA at Space Telescope Science Institute, San Martin Drive, Baltimore, MA, USA
[3]Mathematical Physics Department, University of Torino, Torino, Italy

ABSTRACT

Some non-conventional operating modes, involving both specific detector developments and custom driving, are evaluated in order to define the technical feasibility of direct focal plane coverage and fringe imaging. Special attention has been devoted to resampling techniques, in order to relax the geometrical requirements and/or recover the dynamic range, which appear to be subject to an unsatisfactory trade-off in the case of conventional imaging with standard CCDs.

Application of near infrared detectors to a continuous scanning astrometric survey is discussed, on the basis of the architecture of present-state devices, their performance and near-future technological developments. Several implementations of resampling on CCDs in the visible domain are discussed as well, with reference to the technological requirements involved.

Keywords: interferometry, detector, CCD, GAIA

1. INTRODUCTION

The framework of a medium baseline spaceborn interferometer offers a significant advantage in terms of resolution for several astrometric and astrophysical measures, with respect to the case of a single filled aperture either of equivalent collecting area, or achieving the same diffraction-limited limits.

For a Fizeau assembly, the field of view in which the coherence is mantained can be rather large; requiring that the interference fringes were imaged in detail over a large field of view results in an extremely large pixel count.

Hereafter, we will adopt a framework roughly corresponding to the original GAIA concept (Lindegren & Perryman 1994); many of the conceptual results can be extended easily to different configurations. The basic assumptions are as follows:

- telescope diameter D = 0.5 m;
- baseline B = 2.5 m;
- field of view = $1^\circ \times 1^\circ$;
- fractional bandwidth $\Delta\lambda/\lambda = 25\%$;
- optical scale 11 arcsec/mm;
- rotation period $T_r = 2$ hours.

Some of the previous values are less strictly specified than others: for example, since the optical system must fit inside the bay of an Ariane 5, or a similar class launcher, a much larger baseline is an unlikely possibility, whereas the rotation period can change (in principle) from 2 to 3 hours, with reasonable modifications of the scanning law.

In particular, the optical scale is a physical size trade-off between the minimum resolution element (the 'pixel') and the overall detector assembly, apart from optical considerations. From the above values, assuming direct imaging on the focal plane, the detector assembly size is about 33 × 33 cm. This seems to be acceptable, although challenging, with respect to practical feasibility, thermal constraints and actual mounting in the optical system with the required accuracy and stability.

The feasibility of devices sized ~ 5 cm has been demonstrated (DALSA MegaChip, 5000^2 class, 6.2 cm/side, 12 μm pixels); the focal plane assembly could be implemented by a mosaic of 5×5 or 6×6 chips of 5–6 cm/side.

Requirements for a large field of view and some sort of scanning law are implicit in the all sky survey baseline; anyway, the minimum acceptable field of view deduced in the original concept is 3 times smaller than the above value. Although some oversizing is desirable, to increase the data redundancy and consequently the solutions accuracy, there is possibly margin for trade-offs. This would have significant consequences on all of the system specifications (optics and metrology, detector and detection strategy, attitude monitoring and control).

Hereafter, we will often refer to the along scan direction as the X axis, and to the across scan direction as the Y axis. Due to the scanning operations, the field of view is defined basically by the Y size, whereas the X size defines the exposure time, by integration through the entire field of view; since the rotation speed is $V_r = 360^\circ/T_r = 180$ arcsec/s, the integration time is $T_i = 1^\circ/V_r = 20$ s.

The focal plane assembly physical size is not strictly required to be equal in both directions; therefore, this could be another trade-off possibility, in order to gain design margins, separately for the actual field of view and exposure time.

The single aperture point spread function is the overall envelope of the interference fringes; the pattern can be defined by the Airy disk, sized $\Theta_A = 2.44\ \lambda_{\rm eff}/D$, and the Young period $\Theta_Y = \lambda_{\rm eff}/2B$. Thus, the intrinsic X and Y image resolutions are very different. This leads naturally to a requirement for asymmetric detectors.

We will evaluate at first the involvements of operations in the visible domain, namely $\lambda_{\rm eff} \simeq 550$ nm. The significant figures are $\Theta_A = 554$ mas and $\Theta_Y = 45$ mas, i.e., 50.4 μm and 4.1 μm, respectively. Assuming that the fringe period is sampled over 4 pixels, the along scan pixel size results $\simeq 11$ mas $= 1$ μm; a requirement for optimum sampling would most likely specify 8 pixels/period, or ~ 1 μm size. The across scan size, defined by 2 pixels sampling the central Airy peak, is $\simeq 277$ mas $\simeq 25$ μm.

The sampling strategy will be discussed in more detail in Section 6.. The operating mode is supposed to be some variant of drift scan, or time delay integration, using CCD detectors. Full coverage of the $1^\circ \times 1^\circ$ focal plane assembly at this resolution requires up to 10^9 pixels (3×10^5 linear). The on-chip drift speed, in order to match the scan speed, implies that the time for a CCD row read out must not be larger than that required by a sky point to cross a pixel, because on this time elapse the whole charge pattern is to be advanced by one step. in our framework, this amounts to 61 μs.

Assuming a 1 μs/pixel processing time, a very high number of analog outputs are required in order to read at most 61 pixels on each, thus matching the scan speed. In order to alleviate the requirements on pixel size and count, the adoption of modulation grids has been proposed, but this still features its technical complexity and accuracy requirements, besides resulting in reduced performance compared with the case of direct fringe imaging.

2. PIXEL SIZE

The production of optical detectors with pixel size on the order of 1 μm involves two distinct problems, the former related to the technological implementation itself, and the latter to the achievable performance.

2.1. Technology

In order to estimate the limiting geometry allowed by present state (or near future) electronics technology, the process specifications of some recent devices (mainly high performance microprocessors) have been collected, and are shown in Table 1. Previous data were reported in advertising, press releases and electronics publications, e.g., BYTE; a fully detailed description of all the production processes involved was not available, and it is anyway beyond the scope of this document.

It appears that sub-micron geometry is widely used in mass VLSI (Very Large Scale Integration) production, on several devices, from different manufacturers, on a scale of 0.5–0.7 μm. The trend toward smaller and smaller size is expected to continue, due to the advantages in terms of increased operating speed, reduced power consumption and higher integration (more gates per device and/or smaller devices).

VLSI geometry cannot be reduced indefinitely, because of technical and physical limits; the former is related to the photolithographic process (electron beam technology is expected to reach ~ 100 nm, X-ray photolithography is still beyond production level), whereas the latter is due to the material breakdown (5 V across a 0.5 μm gap build up an electric field of 10 MV/m).

Device	Manufacturer	Geometry [μm]
Pentium P54C	Intel	0.6
68040V	Motorola	0.5
TMS 320C80	Texas Instruments	0.5
Alpha 21064-AA	DEC	0.75 (1992)
Alpha AXP21066	DEC	0.68 (1994)
PowerPC 601	IBM/Apple/ Motorola	0.65 (1993)
R4400PC	MIPS	0.6
Orion R4600	IDT/Toshiba	0.64
PA-RISC 7200	Hewlett Packard	0.55 (1994)
XC5000	Xilinx	0.6

Table 1: VLSI geometry highlights

The IBM Thomas J. Watson Laboratories, in mid 1992, produced several samples of working devices using transistor implemented on 100 nm geometry, by electron beam photolithography (Andrews 1992). *Contrarily to some expectations, quantum effects such as electron tunnelling did not reach levels of significant disturbances to the behaviour of the digital circuit.* Recent IBM achievements can be seen from its World Wide Web page, at address http://www.ibm.com; in particular, from a press release dated 8 February 1995, process CMOS 5S, featuring 0.36 μm channel length, is announced for production in first quarter, 1996. In the next few years, ~ 300 nm geometry should reasonably be a mature and widespread production technology, allowing also for 1–1.5 μm pixel CCD production.

It is worth noting that, although the value specifying the VLSI geometry is only approximately the minimum resolution achievable by a given production process, it is practically the gate size of a MOS structure, usually part of a transistor, but roughly corresponding as well to a CCD electrode.

Most of the devices quoted in Table 1 are produced by means of 3 or 4 metal layer processes, which is suitable only for back-illuminated CCDs; polysilicon gate technology, required for front-illuminated devices, usually uses slightly larger geometry. The production of CCDs featuring pixel size on the order of 1-2 μm, on the basis of a standard 3 electrodes design, seems to be feasible, in the near future, with a reasonable development effort.

2.2. Dynamics

The bucket capacity achievable by a MOS structure, such as a CCD, can be estimated from the performance of present devices, since the basic monocrystal lattice, doping level and sustainable electric field across the silicon dioxide are rather close to the physical limits, and consequently any dramatic improvement seems to be unlikely.

In Table 2 the specifications of the CCD05-30 series, from EEV (GB), are reported. On a 22.5 μm square pixel ge-

CCD05-30 specifications:	(1990)
Pixel pitch	$22.5 \times 22.5\mu$m
Bucket capacity (close to saturation)	3×10^5 e^- min
Dark current (room temperature)	2×10^4 e^- typ
Charge transfer efficiency	99.9995% min
Line transfer time	6.0 μs min
Maximum readout frequency (low noise output)	5.0 MHz min
Readout noise at 140 K (dual slope integration)	10 e^-/pixel rms (10 μs period)
Clocks levels	0–10 V typ

Table 2: CCD05-30 Series Specifications (EEV Ltd)

ometry, the amount of charge which can be allocated to a pixel, close to saturation, is of 3×10^5 e^-, corresponding to a charge density slightly below $\rho_e \simeq 600$ e^-/pixel. Actually, the *linear* range (depending anyway on the acceptable non-linearity threshold) can be considered on the order of 50% the total capacity: $\rho_e \simeq 300$ e^-/pixel. This is consistent with the specifications of a large class of modern devices. The actual (linear) dynamic range depends not only on bucket capacity, but on the readout noise (RON) as well, which in turn depends both on the device performance and on the operating conditions. Higher temperature operations, as well as faster readout, are usually associated with a higher noise budget.

The framework of a long operating life, wide field of view mission, as envisaged, strongly suggests the adoption of passive cooling (allowing for comparably high operating temperature, ~ -50 C) and of fast readout speed. RON performance on slow scan, cooled CCD instruments has reached 2–3 e^- level. Nonetheless, due to the temperature and speed conditions, we will adopt a provisional value RON = 10 e^-. For comparison, the CCD05-30 series devices feature RON $\simeq 6$ e^- on a relatively slow 20 μs/pixel readout using double correlated integration.

On such figures, the pixel size obtained by the sampling requirements of the fringe pattern (1×25 μm) implies a linear bucket capacity of 7500 e^-, corresponding to a linear range of 750, i.e., 7.18 mag. This can be sufficient for some applications, but it is not adequate for observations featuring large dynamic range requirements.

3. WAVELENGTH OPTIONS

The specifications deduced in the introduction assumed operations at $\lambda_{\text{eff}} = 550$ nm; at longer wavelength the fringe pattern size increases, and consequently the values obtained for pixel size, dynamics and pixel count are modified accordingly. The size of the fringe pattern for a few different wavelengths, ranging from the visual to the near infrared, are listed in Table 3, whereas the corresponding pixel size (assuming four pixels fitting in the Young's period and two in the Airy disk) and overall pixel count is presented in Table 4.

The central wavelength has been chosen in such a way as to match the Johnson band system (V, R, I, J and K), whereas the bandwidth has been modified in order to mantain the same relative bandwidth $\Delta\lambda/\lambda \simeq 27\%$, to achieve the same multichromatic fringe visibility.

λ [μm]	$\Delta\lambda$ [μm]	Θ Young [mas]	Θ Airy [mas]
0.55	0.15	45.38	553.61
0.70	0.19	57.75	704.60
0.88	0.24	72.61	885.78
1.25	0.34	103.13	1258.22
2.20	0.60	181.51	2214.46

Table 3: Fringe pattern size versus wavelength

λ [μm]	X size [μm]	Y size [μm]	X count	Y count
0.55	1.03	25.16	317333	13005
0.70	1.31	32.03	249333	10219
0.88	1.65	40.26	198333	8128
1.25	2.34	57.19	139626	5722
2.20	4.13	100.66	79333	3251

Table 4: Pixel size and count versus wavelength

Astrometric accuracy can be degraded by a wavelength increase, since the fringe period (proportional to λ) is also an index of the positional accuracy in the centroid determination. A more accurate estimator is the central fringe rms width.

The basic expression defining the limiting accuracy achievable by an ideal (or nearly-ideal) system, however, depends both on wavelength and signal to noise ratio (through the appropriate factors): $\sigma \sim \lambda/\text{SNR}$. Since the fractional bandwidth is increased proportionally with the effective wavelength, in our framework, the degradation is not dramatic, because of the higher integrated flux and consequent SNR enhancement for photon-limited conditions. Anyway, the budget also depends onto the flux density versus wavelength (Φ), i.e., the spectral type of the objects: SNR $\sim \sqrt{\Phi(\lambda_{\text{eff}}) \times \Delta\lambda}$.

It seems to be reasonable that small changes in λ_{eff} would not substantially modify the list of addressable astrophysical subjects; the scientific involvements of the operating band choice should be explicitly addressed, but this is beyond the scope of this document.

The wavelength increase also relaxes some requirements on the system control accuracy, through the optical path difference OPD = $\lambda^2/\Delta\lambda$. In our assumption of constant relative bandpass, the error margin on OPD increases linearly with the wavelength:

$$\frac{\text{OPD}(1)}{\text{OPD}(2)} = \frac{\lambda_1^2/\Delta\lambda_1}{\lambda_2^2/\Delta\lambda_2} = \frac{\lambda_1}{\lambda_2}$$

The R and I bands attractiveness is due to the significant relaxation in the detector requirements, as shown in Table 4, without the need to relinquish the well established CCD structure and operations. In Section 5. we will describe the near infrared option in more detail. Operations at longer wavelength also result, through the increase in the pixel drift time, on relaxed requirements on analog outputs count and external electronics speed. The I band is particularly interesting, since the pixel size can be increased by 60% and the pixel count is reduced by 37%, in the along scan direction, and might constitute an acceptable trade-off.

4. ACROSS SCAN RESOLUTION

The required positional information is mainly contained in the along scan (X) distribution of the light from any object, accordingly to the fringe pattern obtained from the 'white-light' interferometer; the across scan distribution, shaped accordingly to the Airy disk section, basically, is much less relevant to the measure. The centroid of the fringe pattern provides the high accuracy X position of any given target; its Y position is most effectively achieved in subsequent orbits, in which the scan will cross the same area at different angles.

Spatial resolution on Y is mainly aimed at resolving the confusion among close targets; moreover, it provides informations on the satellite attitude, in addition to the data from the incoherent field of view, useful for full post-hoc reconstruction. Nonetheless, it is possible to trade off the Y resolution, sizing the pixel to values larger than those deduced from the Airy disk sampling requirements, in order to achieve a sufficient silicon area, and consequently dynamic range.

Assuming a Y resolution of 2.2 arcsec, the linear range for the resulting 1×200 μm pixel is 9.45 mag (bucket capacity $\simeq 6 \times 10^4$ e^-, RON $\simeq 10$ e^-, as above). With this strategy, there is a gain of more than 2 mag with respect to the dynamic range corresponding to pixels sized for good sampling in both directions.

The logical Y format of the 1° field of view is 1640 pixels; it could be split onto 6 strips of 275 pixels wide CCDs, with physical size 5.5 cm and logical X format 55000 pixels (total count: 1.51×10^7 elements). The overall pixel count is 5.44×10^8 pixels. Both spatial resolution and dynamic range specifications, as deduced from the detector requirements in our framework, seem to be achievable by near future developments of present state technology.

It is worth investigating the feasibility of additional approaches, in order to provide a set of backup solutions in case the ultimate requirements, for any reasons, could not be met, thus defining reasonable design margins. We shall address the subject in the following.

Depending on the chosen wavelength, the across scan undersampling can be significantly reduced: in I, for example, due to the X pixel size of 1.65 μm, the Y size can be reduced to 120 μm, i.e., 1.3 arcsec, mantaining bucket capacity $\simeq 6 \times 10^4$ e^-, corresponding to a linear dynamic range of about 9.45 mag. This results in a reduced Y confusion limit and improved intrinsic attitude determination.

The detector format is in this case $460 \times 3.33 \times 10^4$. The pixel drift time is 101 μs, thus 5 output lines per chip are sufficient to serve the time delay integration operations. The data rate becomes 9.2 MB/s per device, or 330 MB/s from the whole focal plane assembly. Conversely, with the same 2.2 arcsec Y pixel size, the bucket capacity and dynamic range are 10^5 e^- and 10 mag, respectively, on a device logical format of $275 \times 3.33 \times 10^4$ (9×10^6 pixels total), and a 6×6 chips focal plane assembly.

5. NEAR INFRARED OPERATIONS

As seen in Table 4, the pixel format is further relaxed in the near IR. Assuming similar charge storage capability for near infrared detectors as for CCDs, J band still requires some Y pixel oversizing (80 μm versus 57 μm), whereas in K band the pixel size deduced from sampling criteria allows for as much as 1.2×10^5 e^- bucket capacity. The K detector features logical format $550 \times 1.375 \times 10^4$ on 4×100 μm pixels. In J band, a $700 \times 2.35 \times 10^4$ detector format, on a pixel size 2.34×80 μm, can be assumed.

In classical CCD drift scan observations, the charge pattern built onto the detector is transferred in step with the image, which drifts over the device at a speed defined by declination (on ground) or rotation period (for a spinning satellite). The direct voltage readout architecture of most near infrared detectors, which features individual addressing of every cell, do not allow charge transfer between different pixels, thus preventing the application of the above mechanism. Anyway, it is still possible to perform integrating observations of a slice of the sky drifting along the detector field of view, by means of a different approach taking advantage of the high operating speed of the direct voltage readout structure.

Direct voltage readout devices are often used in co-adding mode, above all in the high background conditions of medium IR observations: since the single exposure time is limited to a few ms, a significant exposure is built up by several hundreds (or thousands) of subsequent frames; their sum is performed by the processing electronics, thus extending in the digital domain the signal integration. This capability can be exploited in a scanning instrument as well.

Basically, a very short exposure is a still image, provided the integration time is shorter than the elapse required by the drift to move the image itself of a detectable (i.e., comparable with the pixel size) amount. Two subsequent frames are images shifted by an amount corresponding to the sky drift in the mean time; on a frame period Δt, the interframe drift is related to the rotation speed $D = V_r \times \Delta t$. The field of view drift must be smaller than the desired spatial resolution, in order to avoid significant blurring: this results in a severe constraint on the readout speed. To achieve $D \leq 10$ mas, the frame period must be $\Delta t = D/V_r \leq 60$ μs.

Although readout cycles of ~ 1 μs/pixel have been demonstrated, and multiple outputs per device are commonplace, it is anyway unlikely that the whole detector could be processed in due time; thus, the adoption of an input catalog, in order to select and read only the relevant portion of the array, seems to be mandatory.

Assuming the frame period as corresponding to a one pixel drift, each subsequent frame is to be added to the logical buffer with a one unit shift in the along scan direction; the last logical row, having received its signal contributions from each physical row all along the transit over the device, is dismissed, whereas the first logical row starts integrating from the values of the physical row just entering the field of view. The net result is that the output data include the contributions of all of the pixels along one column of the device, as in the case of CCD drift scan, although at the cost of a much higher computational load and many readout operations; besides, the dynamics limit toward bright objects is extended significantly, due to the short single exposure time.

It is to be noted that no image recognition mechanism is required, since the geometry defines the inter-frame elapse on the basis of the drift speed; some requirements on the stability of the satellite rotation (speed jitter and drift) can be inferred, but they are basically the same as for the CCD case, since they just relate to image quality conservation on a time scale of seconds, or fractions, apart from the relaxation due to the wavelength increase. Another drawback of near infrared detectors is in their more stringent cryogenics requirements: charge carriers are stimulated by lower energy photons or phonons, thus photocurrent is significant at lower temperature. Consequently, it is less likely that they could be efficiently operated on passive cooling.

6. SAMPLING OPTIONS

In the above, the pixel size was deduced by fitting 4 elements in the Young period of the fringe pattern. This is the simplest application of the Nyquist-Shannon criterium: the fringe distribution features period Θ_Y, with maximum swing $\Delta F = F_{\max} - F_{\min}$ over a distance $\Theta_Y/2$; the sample separation, i.e., the pixel size, should at most reach $\Theta_Y/4$. Actually, the variation rate of the distribution is higher than $\Delta F/(\Theta_Y/2)$ in a small range around the points of inflection, thus in strict sense a still higher sampling rate would be required for complete harmonic analysis (Schwartz 1990). It is common practice to trade off the finest details for a larger field of view (and the principle is mostly compelling in our framework); in standard imaging, it is usually acceptable a setup placing two pixels into the core of the point spread function, i.e., the central peak of the star.

For any object, an interferometer provides a set of fringes instead of a single peak; they contribute significant photometric and astrometric information, due to their strong self-correlation. The fringe pattern is reasonably constant in the rotation period (accordingly to the stringent system stability requirements) and can be calibrated on brighter objects on a rather short time scale. Moreover, several independent samples of the whole fringe pattern are obtained, at least because the field of view is fractioned onto several devices for practical implementation reasons. Every individual detector onto the drift path of a given object can, in principle, acquire an independent measure of its fringe pattern.

Each sample may have, in general, a different phase with respect to the sampling pixel array, i.e., the relative position of pixel sequence and fringes, constant over the observation period corresponding to the transit onto a device, is potentially different between different chips. On the basis of these considerations, we evaluate the feasibility of relaxing the sampling requirements beyond the naïve Nyquist-Shannon formulation, i.e., allowing for a pixel size larger than $S_{NS} = \Theta_Y/4$. Since the detection element, integrating all over its surface, averages the variations on scale smaller than its size, a pixel larger than fringe semi-period could be expected to smear out the difference between maximum and minimum, thus destroying the required information.

Actually, a pixel size $S_{1/2} \simeq \Theta_Y/2$ matches the pixel sequence with the fringe maxima and minima, alternatively, when the phase is properly selected; this suggests that the period information could be preserved. As a mathematical hint, assuming a simplified approximation to the fringe pattern $f(x) = A(1 + \cos 2\pi x/K)$, and integrating over pixels sized K/2, assumed to have rectangular transfer function, and phase x_0, the pixel values still feature a significant degree of modulation, depending on the phase, as from Eq. (1).

$$F_n = \int_{x_0+nK/2}^{x_0+(n+1)K/2} f(x)dx = \qquad (1)$$

$$(AK/2) + (-1)^{n+1}(AK/2\pi)\sin(2\pi x_0/K)$$

Relaxing by a factor of two the detector geometry, the device production, performance and management problems (pixel size, dynamics, data processing) are dramatically modified. The resampling strategy, based on using several correlated samples in order to recover sub-pixel resolution, will be discussed in more detail hereafter; some implementation possibilities will then be outlined. Resampling involves a separation between the concepts of physical pixel size and required resolution, which can be considered, in a sense, the logical pixel size.

6.1. Simulated Samples

We consider the simplified, noiseless, one-dimensional case of two rectangular slits of infinite length, width 55 cm, and center separation 2.45 m. The passband, centered at $\lambda_{\rm eff} = 550$ nm, is Gaussian-shaped, with rms bandwidth $\Delta\lambda = 150$ nm. The resulting fringe pattern and some samples, with different relative phase with respect to the pixel array, and pixel size 10 mas, 20 mas are shown in Fig. 1.

Figure 1: A few samples of the fringe pattern onto an undersampling pixel array, at different phase.

The pixels are assumed as having step response, i.e., no crosstalk and uniform sensitivity. A set of samples (i.e., values from a given pixel sequence) is obtained from the fringe distribution, each shifted by a fixed resampling step $S_R = 1$ mas with respect to the previous one. A sample is the result of the integration, for a given time, of the fringe pattern onto the pixel sequence, with a fixed position phase; the relative position of the samples is known, depending only on the driving system, whereas the actual fringe position is the desired value. The 10 mas pixel case needs (at least) 10 independent samples, whereas 20 of them are required for the 20 mas case, in order to attempt a reconstruction on a 1 mas step (logical pixel size).

The used estimators of the mismatch between the theoretical fringe pattern and the individual samples, or the reconstructed (hereafter, 'resampled') distribution, are the rms width of the central fringe and it visibility, defined as the ratio between its peak value and the first minimum. The ideal fringe pattern features rms width 9.098 mas and visibility 7.08; the values from the single samples are extremely variable, depending on the relative phase between the pixels and the fringes in each of them.

A possible approach may be the selection of the individual samples accordingly to their quality, estimated by fringe rms width, visibility or other estimator. Such a data selection results in the choice of the observation fraction featuring the best match of the pixel sequence versus the fringe pattern. This would lead to a reduction of the effective on-target exposure, and consequent SNR and performance degradation. In order to exploit all of the exposure time, the whole set of samples must be somehow used.

Two approaches have been taken for fringe reconstruction from the undersampled data: total average and floating window. In both cases, the expected result is an estimate of the ideal value of the fringe distribution on a step corresponding to the position (or phase) difference between samples. The total average assigns to each logical pixel the mean value of every physical pixel, in the sample set, including it; the floating window approach attributes to each logical pixel the mean value of all sample pixels centered on it. In a sense, the latter is a 'best guess' value from the whole data set.

Figure 2: Results of fringe pattern reconstruction from the individual samples, through different algorithms, with several pixel sizes.

The results of both algorithms are shown in Fig. 2, for physical pixel sizes ranging from 10 mas to 40 mas. It should be noted that resampling, even in extremely adverse cases, as for 30 and 40 mas pixels, still recovers some of the fringe structure. However, the practical interest lies in the range between 10 mas and 20 mas pixel sizes, roughly corresponding to the geometric range 1 to 2 μm. The analysis, restricted to the central fringe of the theoretical pattern and of both resampling options, of fringe visibility and rms width, gives the results summarized in Table 5.

	Visibility	rms Width [mas]
Ideal pattern	7.08	9.10
Total average:		
10 mas pixels	4.98 (70%)	9.89 (+9%)
20 mas pixels	2.63 (37%)	12.25 (+35%)
Floating window:		
10 mas pixels	5.82 (82%)	9.59 (+5%)
20 mas pixels	3.75 (53%)	10.82 (+19%)

Table 5: Ideal fringe pattern and resampling results

The floating window approach seems to be less sensitive to the pixel size increase than the total average; in both cases, fringe visibility degrades faster than rms width. This resampling scheme evaluation is pessimistic, since it operates only on the central fringe; this might compensate for neglecting noise. The basic concept could be extended to include finite aperture, shot and read-out noise, and smoother pixel transfer function, taking into account non-uniform sensitivity and crosstalk. Each term will add some performance degradation, not only in the measured data, but in the fringe pattern itself (e.g., shot noise); noise terms become negligible for bright objects.

All in all, the significant result is that, apart from the detailed computation depending on the actual figures assumed, there is a whole range of possible detector configurations, departing from the 'ideal sampling' case, in which a trade-off between geometric requirements and measure performance is possible.

7. IMPLEMENTATION OF RESAMPLING

Obtainig a distinct data set from every device would limit the number of samples to just a few units. It is possible to get a larger amount of independent samples, each with lower integration time, by modification of the detector setup, their internal structure, and/or their management.

The approach is different in the case of CCDs and of IR devices: the former requires fractioning the time delay integration onto each chip, whereas direct voltage readout devices mainly require modifications in the data flow processing, without need for specific device structure modifications. The exposure subdivision onto each device also results in relaxing the attitude requirements: since each sample is performed on a smaller elapse of time t_0, only system stability on a comparable time scale is required, instead of over the full elapse of the drift onto a device (about 3 s).

7.1. Phase Relation Between Detectors

The relative phase of the detectors is controlled by the timing system, without the need for particular mounting accuracy. The whole focal plane assembly configuration can be calibrated and monitored on bright objects on short time scale. The satellite rotation (speed $V_r = 180$ arcsec/s) provides a relation between time and position in the focal plane assembly. The gap G_c between chips is covered by the drift in a fixed amount of time $T_c = G_c/V_r$. The phase of the pixel patterns on two neighbouring devices depends both on the gap and on the timing system driving the CCD clocks, since, through proper placement in time of their switching edges, the pixel boundaries can be positioned at will on the drifting image.

Two chips separated by a gap whose size equals an exact multiple of the pixel size features null relative phase: provided they were clocked synchronously, the image on both will be matched to the pixel pattern in the same way. By tuning the delay between the clocking sequences of the devices, a phase shift is introduced, changing the relative pixel positions. The phase can be tuned as well in case of a generic gap, not equivalent to an integer number of pixels. Thus, on a strip of detectors, it is possible to operate them in such a way to obtain a set of samples with a well defined phase relation. The timing resolution required for phase management at $\Delta\phi = 1$ mas is $\tau = \Delta\phi/V_r = 5.6$ μs; the timing accuracy and resolution of many present state electronic devices and modules is more than adequate to cope with such a value.

7.2. Near Infrared Resampling

The coadding mode described in Section 5. is based on the sum of subsequent frames with a logical shift of one pixel, assuming that the frame period were tuned to match this amount of displacement through the drift speed.

The simplest resampling implementation can be described in terms of the separation between the concepts of physical and logical pixel size. The frame period Δt can be tuned to the logical pixel size, smaller than the physical size; consequently, two subsequent frames are shifted by less than one (physical) pixel, and their values are to be remapped onto an higher resolution array in memory for coadding. For example, on a logical to physical pixel size ratio of 1/2, every value obtained from a given physical pixel must be attributed to two adjacent logical pixels; the resampled frames can then be coadded (with one logical pixel shift) to get the full-scan exposure.

The frame period deduced is challenging; the requirement can be significantly alleviated by increasing the rotation period. Speed requirements are relaxed also by means of selective readout: if the position of the target is known, it is possible to read only the relevant part of the detector, ignoring the remaining pixels. Such an approach is potentially affected by crowding limits, due to the finite pixel readout time (~ 1 μs, with present technologies) and consequent limited extension of the detector fraction which can be read in the frame period. The target position can be deduced either by an input catalog, or from the incoherent focal plane assembly detector data.

7.3. Visible Detectors

The visible detectors are supposed to be CCDs operated in time delay integration. Each device provides a sample of the targets, integrated all over its surface. Timing must be mantained within a certain accuracy on the short term, in order to avoid image blurring. The transit time of any object onto a detector, as deduced, is on the order of 3 s, whereas the pixel transit time is on the order of 60 μs.

In order to obtain several samples from a detector, it is necessary to split the integration over the device, separating the distinct data sets. The classic approach on a ground-based camera would simply be closing the shutter, reading the detector and starting another integration; this is not practical for a spaceborn instrument, because of the inconvenience of additional mechanical parts, subject to failure, and of the loss of observing efficiency. The fractionization can be implemented in several ways, all corresponding to the requirement of mantaining the target onto a defined part of the charge pattern for a limited time. Hereafter some of them are described.

7.3.1. Columns overlapping

Since the Y-undersampled fringe pattern is contained into a single pixel column, the exposure time can be limited by a geometry switching columns in the star drift path. This can be implemented by a displacement of the CCD columns, at fixed steps; if the fold spacing corresponds to 1/5 of the total device length ($t_0 = 0.6$ s integration, for a 2 arcmin drift), then 5 samples are obtained from the detector. A sketch of the principle is shown in Fig. 3 (top).

Figure 3: Some practical implementation concepts of on-chip resampling.

The drift path of a target point crosses several columns, remaining on each for an elapse t_0, then switching to the next. Each column thus performs an exposure of duration t_0 on the target. On the exit edge of the device, the 5 samples of the fringe pattern are collected from 5 neighbouring columns. To avoid partial overlap in case of objects close to the column borders, it is possible to dislocate the column structure of an amount larger than one pixel, thus allowing for complete separation of the partial images. The drawback of this approach lies in the across scan confusion limit: assuming Y pixel size of 2 arcsec, any pair of objects closer than 10 arcsec in Y, and ~ 1 arcsec (size of the Airy disk) in X, will get partially overlapped samples. Confusion limit worsens for a larger separation between the steps, e.g., a 1.5 pixel dislocation for 5 integration zones results in an on-chip space occupation of 15 arcsec.

The practical implementation of this approach requires custom design modifications of the production masks for the channel stop diffusions; since this is the lowest resolution layer, this should be achievable without particular problems. The transition between the different column folds cannot be shorter than 1 pixel; for practical implementation reasons, it could be as large as a few pixels, which is nonetheless negligible with respect to the sample exposure.

7.3.2. Forward clocking

The partial images separation can be operated along the columns, instead of among them, by proper clocking, as described in the following. During each time delay integration (of duration t_0), the charge pattern is transferred at drift speed along the columns; then, it can be transferred forward at higher speed, in order to place the target image onto a clear zone behind the present partial image. In order to get non-overlapping fringe patterns, the shift between partial images must be of the order of the Airy disk, i.e., $\sim$ 100 pixels.

After 5 such sequences, 5 samples of the fringe pattern, travelling one after the other along one column, reach the output side, and are read out in sequence (Fig. 3, middle). In this case, there is no confusion in the across scan direction, but only along scan; assuming the partial images were separated by $\sim$ 2 arcsec (Airy disk $\sim$ 1 arcsec), their sequence will occupy about 15 arcsec along scan, but only 2 arcsec across scan. Since the imaging zone of the CCD is logically stretched by the forward clocking process, the mean output readout speed must be higher than the drift speed; in order to match them, an independently controlled CCD area, similar to the storage area of several consumer CCDs, must be implemented before the output register.

7.3.3. Chip tilt

A simple practical implementation can be achieved by placement of the detectors with a small inclination angle with respect to the drift path (Fig. 3, bottom). The constant charge transfer speed, in this case, must have a small mismatch with respect to the drift speed, such that the image changes its phase in subsequent device regions, in which it is also contained in different columns. Thus the partial images corresponding to a given phase are effectively separated.

The confusion is minimal, since a close object will draw a parallel strip, resolved if displaced by more than one pixel in Y and $\sim$ 1 arcsec (Airy disk) in X. The device structure is a regular 2D array of asymmetric, rectangular pixels, clocked at a fixed rate rather close to the drift speed. A side advantage of this approach is that higher information on the target Y position is recovered from its track onto the detector.

7.4. Dynamics Recovery

Since resampling fractionates the integration on every chip, the exposure time of every partial image is reduced by some factor, corresponding to the number of samples obtained from each device: $N_s \sim 5$. This results in a relaxation of the bucket capacity requirements, and correspondingly of the pixel size.

The across scan undersampling can be reduced by N_s, from the $\sim$ 2 arcsec mentioned in Section 4. to about 0.44 arcsec. In such a case, proper sampling in Y direction is recovered in band I, whereas in bands V and R the Airy disk is at least extended over more than one pixel. The confusion limit of the various resampling implementations is then correspondingly reduced. Besides, a reasonable across scan resolution could be very appealing if fringe dispersion were used, as suggested by Noordam 1995.

8. SYSTEM CONSIDERATIONS

From the results of the previous sections, it appears that, in a wavelength range 550-880 nm (V, R and I bands), a CCD with pixel size $\sim 1 - 1.5 \times 40$ μm, or $\sim 11 - 17 \times 440$ mas, can be operated accordingly to some of the described resampling strategies (or further variations), in order to achieve fringe imaging with various performance figures with respect to the ideal sampling case. Such a device features a bucket capacity of $\sim 2 \times 10^4$ e$^-$, i.e., 8.25 mag linear range on 10 e$^-$ read-out noise (and for $\sim$ 0.5 s integration/sample). Its logical format is $\sim 3.9 \times 10^4 \times 1400$ pixels. Due to the pixel transit time of 86 μs, and assuming a pixel processing time of $\sim$ 1 μs, 16 analog outputs per chip are required to keep up with the scan rate.

From the read-out noise assumption of $\sim$ 10 e$^-$, related to high temperature, high speed operations, and the estimated bucket capacity, a minimum electronics resolution of 12 bits can be inferred. Any improvement in the noise budget, of course, would result in enanced instrument performance; our estimate is somewhat conservative, on the basis of readout speed and the comparably high operating temperature, and further developments on detectors performance could result in better noise figures. The adoption of an input catalog (or data from non-coherent detector focal plane assembly) would allow to reduce the number of outputs and/or relax the electronics requirements, performing slower, more accurate readout only on the regions around objects of interest and discarding the remaining part of the sky already at the data acquisition level.

Every 86 μs each of the 6×6 devices provides 1400 pixel values sampled on 12 bits, resulting in a raw data rate of 880 MB/s or over 7 Gb/s. Even assuming a telemetry rate of several Mb/s, challenging for the orbit options foreseen for GAIA, the raw data rate exceeds the available channel capacity of 3 orders of magnitude.

The problem is not likely to be solved just by conventional data compression; consequently, actual data processing is to be performed on-board, in order to reduced the hundreds of raw pixel data for each target to a manageable amount.

The processing power required to sustain the raw data flow is large, and cannot be deduced in full detail without some hint on the algorithms; anyway, it does not seem beyond the capabilities of a few units of present generation processors, rated at several hundreds millions of instructions per second. Processing requirements are further enhanced by the resampling strategy, since for every detector several sets of data are to be combined together for each target. An interesting additional processing task could be the on-line non-linearity correction: our previous dynamics evaluations were based on the linear range of the single pixel, which can in fact accumulate a much larger amount of charge before actual saturation.

Moreover, for very bright objects, even in case of charge bleeding from the central fringe (limited to a few nearby pixels), since an high signal is available from the side fringes, it should be possible to use exclusively the latter to recover the required astrometric information. Both corrections should extend the actual measure dynamic range, by a factor 2–3 each, thus adding ~ 2 mag to the actual operating range.

9. CONCLUSIONS

The feasibility of ideal and nearly ideal fringe detection has been evaluated, with particular attention to approaches allowing to relax the design constraints and to cover a range of performance with different detector configurations. The detector geometry, taking into account the present state of photolithography and the foreseeable near future developments, seems to be able to meet the requirements related to the one-dimensional high resolution and large field of view interferometric environment.

The dynamic range can be enhanced to more than 9 mag by trading off the across scan resolution to gain in pixel area; across scan measure is provided by other observing periods, with different scan direction of the same field. Concerning the sampling requirements, relaxed detector configurations can be envisaged, in which the data from several measures, each with sub-Nyquist resolution with respect to the fringe period, are combined in order to recover additional spatial information. The degradation of the fringe resolution is smooth, thus suggesting the availability of design margins in which intermediate performance can be achieved by less than optimal geometry detectors.

Practical implementation of drift scan observation and resampling strategy, in visible and near infrared bands, has been discussed. The near infrared still seems to involve a much higher technical complexity than visible.

By proper implementation of the overall system configuration, detector assembly, operating mode and on-board processing, a number of viable approaches seems to be available, to be selected both on expected scientific performance and technical achievements attainable in the near future.

REFERENCES

Andrews, D., 1992, BYTE, August issue, 28

Lindegren, L., Perryman, M.A.C. 1994, GAIA: Global Astrometric Interferometer for Astrophysics (A Concep for an ESA Cornerstone Mission) Technical report

Noordam, J.E. 1995, ESA SP–379, this volume

Schwartz, M. 1990, Information Transmission, Modulation and Noise - Fourth Edition McGraw-Hill International Editions