

OPTIMAL LINEAR IMAGE COMBINATION

BARNABY ROWE,^{1,2} CHRISTOPHER HIRATA,³ AND JASON RHODES^{1,2}*To appear in ApJ*

ABSTRACT

A simple, yet general, formalism for the optimized linear combination of astrophysical images is constructed and demonstrated. The formalism allows the user to combine multiple undersampled images to provide oversampled output at high precision. The proposed method is general and may be used for any configuration of input pixels and point spread function; it also provides the noise covariance in the output image along with a powerful metric for describing undesired distortion to the image convolution kernel. The method explicitly provides knowledge and control of the inevitable compromise between noise and fidelity in the output image.

We present a first prototype implementation of the method, outlining steps taken to generate an efficient algorithm. This implementation is then put to practical use in reconstructing fully-sampled output images using simulated, undersampled input exposures that are designed to mimic the proposed *Wide Field InfraRed Survey Telescope* (WFIRST). We examine results using randomly rotated and dithered input images, while also assessing better-known “ideal” dither patterns: comparing results we illustrate the use of the method as a survey design tool. Finally, we use the method to test the robustness of linear image combination when subject to practical realities such as missing input pixels and focal plane plate scale variations.

Subject headings: dark energy — *methods:* data analysis — *techniques:* image processing — *techniques:* photometric — weak gravitational lensing

1. INTRODUCTION

Weak gravitational lensing (e.g. Schneider 2006) is a powerful means by which to probe the distribution of matter in the Universe. As such, it has been identified as among the most promising methods by which to constrain the growth rate of matter structure in the Universe, and thereby to test models of evolving dark energy and modifications to general relativity (see, e.g., Albrecht et al. 2006, 2009; Peacock et al. 2006, and references therein). However, weak lensing is arguably the most technically challenging of the cosmological probes from an image analysis standpoint, due to its extreme sensitivity to any systematic errors made when recovering galaxy shape information. Much work has gone into understanding how to accurately recover tiny, coherent gravitational shear signals from large numbers of noisy galaxy images that have been convolved with the instrumental Point Spread Function (PSF) (Heymans et al. 2006; Massey et al. 2007; Bridle et al. 2010; Kitching et al. 2010).

In the analysis of astronomical images, ensuring the adequate spatial sampling of data by pixels of finite size and spacing is a key concern. Ideally, images should be sampled at or above the Nyquist-Shannon sampling rate for the *band limit* set by the optical response of the system (see e.g. Marks 2009), so that the full continuous image can be determined from the discrete pixel samples. Throughout this Paper we will define a single plane wave to be $\propto e^{2\pi i \mathbf{u} \cdot \mathbf{r}}$, so that \mathbf{u} has units of cycles per arc second. In general, if an image contains only Fourier modes whose spatial frequency \mathbf{u} is no larger in magnitude than u_{\max} then the Nyquist criterion demands sample spacing P satisfying $P < 1/(2u_{\max})$. An image sam-

pled more finely than the critical rate $1/(2u_{\max})$ is referred to as *oversampled* (and one sampled at the critical rate is *critically sampled*). Since it is possible to reconstruct the entire original image from oversampled data, operations such as interpolation/regridding, rotation, and translation can be carried out with no pixelization artifacts. This makes oversampled data the preferred input for most precision image analysis applications, including weak lensing.

The ideal location for weak lensing observations is a space-based telescope, where one is free from the blurring effects of the Earth’s atmosphere and can achieve a level of stability of the optics and hence the PSF that is impossible from the ground. Weak lensing is thus a key project for proposed space-based imaging missions such as the *Wide Field InfraRed Survey Telescope* (WFIRST: Blandford et al. 2010; the reference design is based on the *Joint Dark Energy Mission/JDEM-Omega* concept: Gehrels 2010) and *Euclid* (Refregier et al. 2010). However, in both cases, practical design considerations prohibit oversampling at the native pixel scale of the system. The optics of a space telescope deliver a PSF that preserves Fourier modes out to $u_{\max} = D/\lambda$ where D is the outer diameter of the primary mirror and λ is the wavelength of observation; the high spatial frequencies may be suppressed by e.g. charge diffusion, but in most cases u_{\max} is still large. Indeed, preserving the high spatial frequency components of the image is a major reason to choose a space-based platform. If one is to oversample the image at the native pixel scale, then, one is forced to choose very small pixels. However, there are competing considerations that drive one to larger pixels, including (i) the desire for large field of view within engineering or cost constraints on the number of detectors; and (ii) the high read noise of near-infrared detectors, which results in increased photometric errors as light from an object is spread over more pixels. For these reasons, both *JDEM-Omega* and *Euclid* chose pixels larger than $1/(2u_{\max})$. Images from such systems, sampled at the native pixel scale, are *undersampled*. Undersampled im-

browe@caltech.edu; chirata@tapir.caltech.edu; jason.d.rhodes@jpl.nasa.gov

¹ Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109² California Institute of Technology, 1200 E California Boulevard, Pasadena, CA, 91106³ Department of Astrophysics, California Institute of Technology, M/C 350-17, 1216 E California Boulevard, Pasadena, CA, 91106

ages suffer from aliasing, in which each Fourier mode of the observed image contains contributions from several Fourier modes of the original image: in this case the original image cannot be unambiguously reconstructed. Operations such as interpolation, translation by a non-integer number of pixels, etc. are not mathematically possible on undersampled data, which represents a major difficulty for image analysis.

Although proposed space-based dark energy missions are unlikely to oversample their imaging data on a given single image, typical design concepts allow accurate pointing capability (e.g. Gehrels 2010). What is required is thus a means of reconstructing oversampled images from two or more undersampled, dithered images. The simplest technique to improve sampling is the ideal subpixel dither: for example, one may take 4 exposures whose relative offsets are $(0, 0)$, $(\frac{1}{2}, 0)$, $(0, \frac{1}{2})$, and $(\frac{1}{2}, \frac{1}{2})$ pixels (a “ 2×2 ” dither pattern). With the use of these multiple exposures, where the observed image depends on different linear combinations of the aliased modes, one may use linear algebra techniques to separately “solve out” the aliased Fourier modes (Lauer 1999a,b) so long as the native pixel scale $P < 1/u_{\max}$. The linear algebra techniques allow us to analyze any subpixel dither pattern: notable examples include the “ 3×3 ” pattern (9 exposures); subpixel dithers with diagonal grids such as the “ $\sqrt{2} \times \sqrt{2}$ ” pattern [where the relative offsets are $(0, 0)$ and $(\frac{1}{2}, \frac{1}{2})$ pixels] or the “ $\sqrt{5} \times \sqrt{5}$ ” pattern (see Section 5.2); or N exposures with arbitrary offsets. Given any such subpixel dither pattern, one can determine whether the linear system for the Fourier modes of the input image can be solved. The output image from this approach has the same PSF as the input images.

However, it is likely that *WFIRST* – or any wide-angle survey that attempts to fill in gaps between detectors – will end up with a survey pattern that contains large slews (of order a detector width) between each exposure. Under such circumstances, the different input images that are being used to build the final, oversampled output image will be sampled on different grids whose phase cannot be controlled, will have different field distortions, and will have “holes” due to cosmic rays and detector effects. Moreover, the input PSFs will be different even if the telescope optics are perfectly stable, because each observation of a galaxy will be made in a different part of the field. The Fourier space/linear algebra techniques are not well-equipped to handle such situations efficiently. There are stable, commonly-used image combination algorithms such as DRIZZLE (Fruchter & Hook 2002) with mature accompanying scripts (e.g. Koekemoer et al. 2002) that will produce an output image for such inputs. However, these algorithms do not give full control of the PSF in the output image, or even guarantee that this PSF is constant across the image of a galaxy⁴, which should be seen as a prerequisite for high-precision shape measurement for a space weak lensing mission. Recently Fruchter (2011) proposed a new method (IDRIZZLE: iterative DRIZZLE) that tackles some of these issues using an iterative application of DRIZZLE commands. It will be interesting in the future to compare results using this algorithm to those of this Paper.

A third approach to combining images would be to adapt methods from Cosmic Microwave Background (CMB) ex-

periments. CMB experiments typically have only a few pixels in their focal plane (the *Wilkinson Microwave Anisotropy Probe*/WMAP had a total of 10 differencing assemblies: e.g. Jarosik et al. 2011, and references therein) and use a large number of scans across the sky to build up an output map. Mapmaking for CMB experiments is usually done by taking the vector \mathbf{d} of time-ordered data from a small number of detectors, and writing it in the form:

$$\mathbf{d} = \mathbf{M}\mathbf{s} + \boldsymbol{\eta}, \quad (1)$$

where \mathbf{s} is a vector of temperatures in pixels on the sky, $\boldsymbol{\eta}$ is the noise, and \mathbf{M} is the $n_d \times n_p$ mapping matrix (n_d being the number of samples of data and n_p the number of pixels in the output map). In its simplest form \mathbf{M} is a sparse matrix containing $M_{ij} = 1$ if the i th sample is collected when looking at the j th pixel, and 0 otherwise, but many generalizations are possible (e.g. for differential experiments, methods that mask bright or variable sources so they do not contaminate the data processing, etc. see Hinshaw et al. 2007). Taking the known noise covariance matrix $\mathbf{N} = \langle \boldsymbol{\eta}\boldsymbol{\eta}^T \rangle$, one may obtain a least-squares solution for the sky map \mathbf{s} : this is $(\mathbf{M}^T \mathbf{N}^{-1} \mathbf{M})^{-1} \mathbf{M}^T \mathbf{N}^{-1} \mathbf{d}$.

However, this method (as written here and in most of the literature) does not trivially incorporate a variable PSF (beam), requires the output \mathbf{s} to have pixels *much* smaller than the beam, and then requires at least several samples per output pixel – this is a luxury that is not present in optical astronomy with sub-arcsec PSF sizes. This problem can be solved by allowing \mathbf{M} to contain interpolation coefficients, but an implementation of this approach is not yet public, and is typically unnecessary for the arcminute beam sizes commonly seen in CMB experiments.

In this Paper, we take the first steps toward developing a general linear algorithm that attempts to combine the best features of these approaches. As in the Fourier/linear algebra approach (Lauer 1999a,b), we desire a well-controlled final PSF. As is the case for DRIZZLE (Fruchter & Hook 2002), we want to be able to use arbitrarily placed samples if possible. Sometimes, reconstructing a fully-sampled image with a controlled PSF is impossible given the positions of the samples provided; unlike DRIZZLE, we desire a method that alerts the user when this happens. Finally, as in the CMB approach, we aim to do something that minimizes output noise using some variant of the least-squares method.

In Section 2 we derive a linear formalism that can meet these aims, and in Section 3 describe a prototype implementation of the method called IMCOM, which is available freely from the authors on request. In Section 4 we illustrate the process of combining multiple images using IMCOM via a detailed worked example, before turning in Section 5 to examine a set of realistic potential observing scenarios for a space-based dark energy mission such as *WFIRST*. While it should be stressed that the dither scenarios and PSF patterns used in these tests do not necessarily reflect any firm plans for the *WFIRST* mission, they are designed to approximate such a mission in a broad sense. In Section 6 we discuss the computing resources necessary to process imaging data from a dark energy mission using the IMCOM technique, before drawing conclusions in Section 7.

2. LINEAR FORMALISM

The problem is to combine several undersampled input images into a single oversampled image. The input images may

⁴ Special cases such as the 2×2 ideal subpixel dither with identical input PSFs may be exceptions because of special symmetries, e.g. translation by half-pixels. However, our interest is in the combination of images with no such symmetries.

be written as a vector of intensities I_i of length n , where n is the total number of usable pixels. For example, if the inputs are four 128×128 pixel images with no defects, then $n = 4(128)^2$. The inputs are sampled at their pixel centers $\{\mathbf{r}_i\}_{i=1}^n$. We also suppose that the PSF (including jitter, optics, and detector response) at separation \mathbf{s} is $G_i(\mathbf{s})$. That is, the intensities are given by

$$I_i = \int_{\mathbb{R}^2} f(\mathbf{r}') G_i(\mathbf{r}_i - \mathbf{r}') d^2 \mathbf{r}' + \eta_i, \quad (2)$$

where η_i is the noise with $\langle \eta_i \rangle = 0$ and some covariance matrix $N_{ij} = \langle \eta_i \eta_j \rangle$. The formalism can rigorously treat any general noise covariance matrix N_{ij} , although in most cases N_{ij} is close to diagonal (with small off-diagonal correlations due to, e.g., corrections for inter-pixel capacitance, electronic $1/f$ or flicker noise, etc.). The function $f(\mathbf{r}')$ describes the physical image on the sky.

We note here that the PSF $G_i(\mathbf{r})$ is assumed as being fully specified, in advance, at the position centered on each input pixel \mathbf{r}_i , in each contributing exposure individually. This knowledge must be expected to come from either a well-motivated optical model or the image data I_i , or both. Of course, images of point sources (stars) within I_i will be subject to the same degree of undersampling as the rest of the I_i pixels. This issue can be countered if adopting a fully parametric model and fitting for $G_i(\mathbf{r})$ from stellar images (e.g. Ma et al. 2008). However, the use of a non-parametric model that potentially contains defects due to aliasing will undermine the success of any image combination. The problem of recovering $G_i(\mathbf{r})$ from undersampled stellar images is discussed further in Section 7.

The output is to be a synthesized, fully sampled image H_α on a grid of pixel centers \mathbf{R}_α , where $\alpha = 1, \dots, m$. For example, if the synthesized image is to be 256×256 then $m = 256^2$. We wish for H_α to be as close as possible to the target image J_α , which is defined as:

$$J_\alpha \equiv \int_{\mathbb{R}^2} f(\mathbf{r}') \Gamma(\mathbf{R}_\alpha - \mathbf{r}') d^2 \mathbf{r}'. \quad (3)$$

Here Γ is the desired effective PSF of the synthesized image, a free choice for the user. A well-motivated choice for Γ might be $G_i(\mathbf{r})$ itself (if it is approximately constant), preserving the input PSF as much as possible. This is the approach taken in later Sections, but in general Γ may be freely chosen to additionally filter the output image as desired (see, e.g., Section 5.5).

2.1. Notation

Fourier transform pairs are written with Fourier space in units of cycles per arcsecond rather than radians per arcsecond. That is,

$$\tilde{f}(\mathbf{u}) = \mathcal{F}[f(\mathbf{r})](\mathbf{u}) = \int_{\mathbb{R}^2} f(\mathbf{r}) e^{-2\pi i \mathbf{u} \cdot \mathbf{r}} d^2 \mathbf{r} \quad (4)$$

for the forward Fourier transform, and

$$f(\mathbf{r}) = \mathcal{F}^{-1}[\tilde{f}(\mathbf{u})](\mathbf{r}) = \int_{\mathbb{R}^2} \tilde{f}(\mathbf{u}) e^{2\pi i \mathbf{u} \cdot \mathbf{r}} d^2 \mathbf{u} \quad (5)$$

for the inverse Fourier transform.

The Einstein summation convention is *not* used: all summations over indices will be either be written explicitly or, for brevity, as implied matrix multiplications upon matrix and

vector objects in bold face. Latin indices are used for input vectors and in derived objects where reference is made to input pixel locations, and Greek indices are similarly used to refer to output pixel locations.

2.2. Linear Solutions

We have defined the fully-sampled output H_α in relation to the target image J_α of equation (3), to which we intend H_α to be as close an approximation as possible. We then desire H_α to be linearly related to the input images I_i for several reasons. Most importantly, it ensures that the final PSF in H_α is independent of image brightness, so the stars that were not used in the PSF determination are tracers of how well the image synthesis worked. Linearity simplifies the analysis of systematics. Furthermore, it dramatically reduces the space of possible algorithms and allows a systematic search of the space that remains. We further note that some pre-existing algorithms (e.g. DRIZZLE: Fruchter & Hook 2002) are also linear and hence are included in our search space. It should be made clear, however, that detectors usually display some level of non-linear behavior (e.g. non-linear gain), changing the PSF for bright objects. Corrections for this non-linearity are typically applied in image pre-processing at the raw pixel level, and we assume that any such non-linearity will have been duly corrected in I_i before starting to combine images.

A general linear mapping from $I \rightarrow H$ is:

$$H_\alpha = \sum_i T_{\alpha i} I_i, \quad (6)$$

where $T_{\alpha i}$ is an $m \times n$ matrix. We now need to construct an appropriate objective function for the performance of the matrix $T_{\alpha i}$ and extremize it to find the optimal linear transformation we desire.

2.3. Objective Function

We note that for a matrix $T_{\alpha i}$, an error map can be constructed:

$$\begin{aligned} Z_\alpha &\equiv H_\alpha - J_\alpha \\ &= \int_{\mathbb{R}^2} f(\mathbf{r}') L_\alpha(\mathbf{R}_\alpha - \mathbf{r}') d^2 \mathbf{r}' + \sum_i T_{\alpha i} \eta_i \end{aligned}$$

where we have defined the leakage function L_α as

$$L_\alpha(\mathbf{R}_\alpha - \mathbf{r}') \equiv \sum_i T_{\alpha i} G_i(\mathbf{r}_i - \mathbf{r}') - \Gamma(\mathbf{R}_\alpha - \mathbf{r}'). \quad (7)$$

The leakage is simply the difference between the desired PSF Γ and its actual reconstructed counterpart in H_α . We will seek to minimize this difference. Taking equation (7), we may then separate out the deterministic part,

$$\langle Z_\alpha \rangle = \int_{\mathbb{R}^2} f(\mathbf{r}') L_\alpha(\mathbf{r}' - \mathbf{R}_\alpha) d^2 \mathbf{r}', \quad (8)$$

from the stochastic contribution to the residuals as specified by the noise covariance,

$$\Sigma_{\alpha\beta} = \text{Cov}(Z_\alpha, Z_\beta) = \sum_{ij} T_{\alpha i} T_{\beta j} N_{ij}. \quad (9)$$

We will also seek to minimize the noise variance $\Sigma_{\alpha\alpha}$ for each output pixel value H_α . Noisy output images will be those that combine input pixels from I_i with large positive and negative weights in $T_{\alpha i}$. Such solutions will tend to be unstable, which

gives a further practical reason for wishing to minimize $\Sigma_{\alpha\alpha}$ where possible,

This now allows the construction of an objective function W , i.e. a function to be minimized in solving for $T_{\alpha i}$. We will choose a function that is a simple sum over values of α , so that if we wish to we may write $W = \sum_{\alpha} W_{\alpha}$, where W_{α} depends only on $L_{\alpha}(\mathbf{s})$ and $\Sigma_{\alpha\alpha}$. In other words, we choose to optimize each output pixel independently: one minimizes W by minimizing each W_{α} .

This choice is made both for computational simplicity, and also so that the optimization does not mangle good regions of the synthesized image to try to clean up bad regions (e.g. the edges of frames or the occasional intersection of multiple cosmic ray tracks). We desire for W_{α} to be a quadratic function, so that linear algebra methods can be used to solve for $T_{\alpha i}$, and for it to be translation-invariant (i.e. we care only about the normalization and shape of the leakage function L_{α} , not about its position on the sky). The most general such function is

$$W_{\alpha} = U_{\alpha} + \kappa \Sigma_{\alpha\alpha}, \quad (10)$$

where $\kappa > 0$ is a Lagrange multiplier, and we have defined the *leakage objective* U_{α} as

$$U_{\alpha} = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \Upsilon(\mathbf{s}' - \mathbf{s}) L_{\alpha}(\mathbf{s}) L_{\alpha}(\mathbf{s}') d^2\mathbf{s} d^2\mathbf{s}', \quad (11)$$

where $\Upsilon(\mathbf{s})$ is a real symmetric positive-definite kernel, i.e. $\Upsilon(-\mathbf{s}) = \Upsilon(\mathbf{s})$, $\Im \Upsilon(\mathbf{s}) = 0$, and $\Upsilon(\mathbf{s}) > 0$. We note that for the simplest choice of $\Upsilon(\mathbf{s})$, the Dirac delta function $\delta(\mathbf{s})$, this expression becomes

$$U_{\alpha} = \int_{\mathbb{R}^2} [L_{\alpha}(\mathbf{s})]^2 d^2\mathbf{s}. \quad (12)$$

The leakage objective U_{α} is thus, in the case $\Upsilon(\mathbf{s}) = \delta(\mathbf{s})$, a very simple, direct, quadratic measure of the local leakage at the point \mathbf{R}_{α} . Returning to the general case, equation (11) may also be written conveniently in Fourier space:

$$U_{\alpha} = \int_{\mathbb{R}^2} \tilde{\Upsilon}(\mathbf{u}) |\tilde{L}_{\alpha}(\mathbf{u})|^2 d^2\mathbf{u}. \quad (13)$$

It will be seen this expression for U_{α} is computationally convenient.

The role of the Lagrange multiplier κ in equation (10) merits some discussion here. It can be seen that κ may be used to weight the relative importance of the two terms contributing to the objective function W_{α} . These relate to two key properties of the output image:

- The *fidelity* of the ensemble mean H_{α} to the target image of equation (3), on which U_{α} depends.
- The noise variance in the output image, $\Sigma_{\alpha\alpha}$.

To ensure fidelity we naturally wish to emphasize the minimization of U_{α} relative to $\Sigma_{\alpha\alpha}$, implying $\kappa \ll 1$, and vice versa if we wish to suppress noise. Thus the minimization of W_{α} to solve for $T_{\alpha i}$, subject to the freely-chosen Lagrange multiplier constraint κ , gives the user control of the balance between image fidelity and noise at each output pixel H_{α} .

2.4. Minimization of the Objective Function

We want to minimize W_{α} over the space of $T_{\alpha i}$. To do this, we first take the Fourier transform of equation (7):

$$\tilde{L}_{\alpha}(\mathbf{u}) = \left[\sum_i T_{\alpha i} e^{-2\pi i \mathbf{u} \cdot \mathbf{r}_i} \tilde{G}_i^*(\mathbf{u}) - e^{-2\pi i \mathbf{u} \cdot \mathbf{R}_{\alpha}} \tilde{\Gamma}^*(\mathbf{u}) \right] \times e^{2\pi i \mathbf{u} \cdot \mathbf{R}_{\alpha}} \quad (14)$$

The leakage objective is then

$$U_{\alpha} = \int_{\mathbb{R}^2} \left| \sum_i T_{\alpha i} e^{-2\pi i \mathbf{u} \cdot \mathbf{r}_i} \tilde{G}_i^*(\mathbf{u}) - e^{-2\pi i \mathbf{u} \cdot \mathbf{R}_{\alpha}} \tilde{\Gamma}^*(\mathbf{u}) \right|^2 \times \tilde{\Upsilon}(\mathbf{u}) d^2\mathbf{u}. \quad (15)$$

Recalling that $T_{\alpha i}$ is real, and using equations (9) & (10), this expands to give the combined objective function:

$$W_{\alpha} = \sum_{ij} (A_{\alpha ij} + \kappa N_{ij}) T_{\alpha i} T_{\alpha j} + \sum_i B_{\alpha i} T_{\alpha i} + C_{\alpha}, \quad (16)$$

where the coefficients are given by the system matrices:

$$A_{\alpha ij} = \int_{\mathbb{R}^2} \tilde{\Upsilon}(\mathbf{u}) e^{2\pi i \mathbf{u} \cdot (\mathbf{r}_j - \mathbf{r}_i)} \tilde{G}_i^*(\mathbf{u}) \tilde{G}_j(\mathbf{u}) d^2\mathbf{u}, \quad (17)$$

$$B_{\alpha i} = -2 \int_{\mathbb{R}^2} \tilde{\Upsilon}(\mathbf{u}) e^{2\pi i \mathbf{u} \cdot (\mathbf{r}_i - \mathbf{R}_{\alpha})} \tilde{G}_i(\mathbf{u}) \tilde{\Gamma}^*(\mathbf{u}) d^2\mathbf{u}, \quad (18)$$

$$C_{\alpha} = \int_{\mathbb{R}^2} \tilde{\Upsilon}(\mathbf{u}) |\tilde{\Gamma}(\mathbf{u})|^2 d^2\mathbf{u}. \quad (19)$$

Here $A_{\alpha ij}$ is an $n \times n$, positive definite symmetric real matrix, $B_{\alpha i}$ is a real matrix of dimension $m \times n$, and C_{α} is a positive real scalar. (The integrals can be seen to be real since the integrands in $A_{\alpha ij}$, $B_{\alpha i}$, and C_{α} transform to their complex conjugates under change of the integration variable $\mathbf{u} \rightarrow -\mathbf{u}$.) We note that direct expansion of equation (15) would have lead to a Hermitian matrix; however, since $T_{\alpha i}$ is real the contributions to U_{α} from the imaginary, antisymmetric part vanish.

Any function written in the form of equation (16) can be immediately be minimized to give

$$T_{\alpha i} = -\frac{1}{2} \sum_j \left[(\mathbf{A}_{\alpha} + \kappa \mathbf{N})^{-1} \right]_{ij} B_{\alpha j}, \quad (20)$$

where $(\mathbf{A}_{\alpha} + \kappa \mathbf{N})^{-1}$ is the $n \times n$ matrix inverse of $A_{\alpha ij} + \kappa N_{ij}$. In practice it will be necessary to restrict the synthesized image pixel α to only depend on nearby input pixels i so as to reduce the dimensionality of the matrix $A_{\alpha ij}$. If this is done, however, $A_{\alpha ij}$ actually depends on α which is not true if the whole image is used. Since the whole image may have $n \sim 10^7$, this reduction of dimensionality is required in order to make the approach computationally feasible. These issues are discussed further in Section 3.

We note that equation (20) is a general result and can be used in principle for any system of non-ideal dithers, rotated images, and even images with defects (the “lost” pixels are simply removed from the system matrices). Some of these cases will be explored in Sections 4 & 5.

The noise variance in each output image pixel is given by the diagonal terms in the full covariance of equation (9). As discussed in Section 2.3, the noise properties can be adjusted to using the Lagrange multiplier κ . In Appendix A we derive in detail the limiting behavior of output images across the

range of κ values. These results make it clear that, in setting κ , some compromise is desired between the limiting cases of maximum fidelity and minimum noise. In some cases, even if κ is taken to zero (i.e. the optimization is sensitive only to mean output image fidelity and not to noise) the leakage function will not approach zero. In these cases, reconstructing an unbiased synthesized image is *not mathematically possible* using a purely linear combination of the input pixels, and $\lim_{\kappa \rightarrow 0^+} U_\alpha$ represents a fundamental limitation.

The choice of κ may be made globally or from pixel to pixel. Setting a single value of κ is certainly the simplest approach, but it is advantageous to allow the user to specify the properties of their input image in terms of key deliverables such as the noise variance $\Sigma_{\alpha\alpha}$ or leakage objective U_α . One strategy for image combination would therefore be to regard $\Sigma_{\alpha\alpha}$ as fixed (using this to set the value of κ from pixel to pixel) and then obtain the highest-fidelity mean image consistent with this noise value. The opposing strategy would be to set a tolerance on the image fidelity by requiring U_α to be less than some threshold value, and then finding the least noisy reconstruction consistent with this threshold. We now consider these strategies and give more detail about practical issues in the implementation of the algorithm.

3. IMPLEMENTATION

In this Section we describe a prototype implementation of the image combination method presented in Section 2, describing some of the numerical techniques employed and issues raised in finding solutions for $T_{\alpha i}$. We have named this prototype package IMCOM (IMage COMbination), and plan to make it available to the public either in its current version or as part of a more developed software package at a later date. The code is written in FORTRAN 95, supports multi-threading via OPENMP (see, e.g., Chapman, Jost, & Pas 2007), and uses only public software libraries, some of which are available in specific hardware-optimized versions: CFIT-SIO (Pence 1999), FFTW3 (Frigo & Johnson 2005), BLAS (Blackford et al. 2002) & LAPACK (Anderson et al. 1999).

3.1. IMCOM Program Outline and Inputs

The IMCOM code solves for $T_{\alpha i}$ by varying κ on an output pixel-by-output pixel basis as described in Section 2.4, so that $\kappa = \kappa_\alpha$. The user may specify a maximum tolerance value of either the noise variance or leakage objective: these we label $\Sigma_{\alpha\alpha}^{\max}$ and U_α^{\max} , respectively. The algorithm may then be used to solve for κ_α so that $\Sigma_{\alpha\alpha} \leq \Sigma_{\alpha\alpha}^{\max}$ or $U_\alpha \leq U_\alpha^{\max}$, while simultaneously minimizing the corresponding U_α or $\Sigma_{\alpha\alpha}$, respectively. This procedure leaves the user free to specify an appropriate compromise between noise and fidelity for the application in question.

In practice it is desirable to have an iterative solution stop within finite time, and this can be achieved by specifying an acceptable range for the solutions, $\Delta\Sigma_{\alpha\alpha}^{\max}$ and ΔU_α^{\max} . The algorithm can then be instructed to finish iterating once $\Sigma_{\alpha\alpha}^{\max} - \Delta\Sigma_{\alpha\alpha}^{\max} < \Sigma_{\alpha\alpha} \leq \Sigma_{\alpha\alpha}^{\max}$ or $U_\alpha^{\max} - \Delta U_\alpha^{\max} < U_\alpha \leq U_\alpha^{\max}$. This then allows total control over either the noise or fidelity of the output.

Apart from these tolerances on the properties of the output image, there are other choices that the user needs to make before the problem is fully specified. These we refer to as “soft inputs”, since the user has some freedom in the choice of their functional forms or values:

- i) The synthetic convolution kernel $\Gamma(\mathbf{r})$ of the desired

output image.

- ii) The desired output sampling locations \mathbf{R}_α .
- iii) A choice of objective kernel $\Upsilon(\mathbf{r})$ for evaluating U_α . The simplest choice, made throughout this paper, is a Dirac delta function so that $\tilde{\Upsilon}(\mathbf{u}) = 1$.
- iv) The total range of search for solutions that match the input tolerances on leakage or noise: $\kappa_{\min} \leq \kappa_\alpha \leq \kappa_{\max}$.
- v) The maximum allowed number of interval bisections n_{bis} of this κ range in the search for solutions.

“soft inputs”

There are additional, low-level, implementation-specific parameter choices that affect the degree of numerical approximation made when calculating $A_{\alpha ij}$ and $B_{\alpha i}$. These are described in Section 4.2, and the values used are tabulated (along with fixed choices of the soft inputs) in Table 1.

The algorithm also requires input values and functions which are *not* chosen, and are determined by the images in question. These “hard inputs” are assumed to be fully known in advance, and are simply:

- i) The input images arranged into a continuous vector I_i and corresponding input pixel center positions \mathbf{r}_i .
- ii) The PSF at the position of each pixel center $G_i(\mathbf{r})$.
- iii) The noise covariance of the image pixels N_{ij} .

Once all of these inputs are given the problem is fully specified, and the algorithm can begin constructing the optimal linear transformation $T_{\alpha i}$.

3.2. System Matrices from the Fast Fourier Transform

The first significant computational task in the algorithm is the calculation of the system matrices $A_{\alpha ij}$ and $B_{\alpha i}$, and the scalar C_α , given in equations (17)-(19). We assume that $G_i(\mathbf{r})$ and $\Gamma(\mathbf{r})$ will be submitted to the software as fully-sampled, discrete images, although if some sufficiently accurate analytic form is known this could also be used instead.

A numerical approximation to the integrals of equations (17)-(19) might then be calculated directly for each α, i, j element in turn. This approach begins by constructing discrete representations of $\tilde{G}_i(\mathbf{u})$ and $\tilde{\Gamma}(\mathbf{u})$ using the Discrete Fourier transform (DFT: see, e.g., Press et al. 1992) of the input $G_i(\mathbf{r})$ and $\Gamma(\mathbf{r})$, then using these functions to take a direct sum over the integrand. However, a DFT of finite size in truth represents each PSF image as a periodic function in real space, which can cause erroneous successive maxima in the system matrices wherever the combination $(\mathbf{r}_j - \mathbf{r}_i)$ exceeds the dimensions of the original PSF image; the same is true for $(\mathbf{r}_i - \mathbf{R}_\alpha)$ when calculating $B_{\alpha i}$. This problem can be overcome by zero-padding $G_i(\mathbf{r})$ and $\Gamma(\mathbf{r})$ to match the spatial extent of the input \mathbf{r}_i and output \mathbf{R}_α , and then taking the DFT. The Fast Fourier Transform (FFT: e.g., Press et al. 1992) algorithm performs this task highly efficiently.

Yet this schema becomes prohibitively slow in the case of large, well-sampled PSF images. If we assume that the PSF images are better sampled than our input I_i by a factor n_{PSF} along each dimension, then the FFT of the zero-padded PSF images requires $O[n_{\text{PSF}}^2 n \log(n_{\text{PSF}}^2 n)]$ calculations. That cost may be borne, but going on to calculate $A_{\alpha ij}$

then requires a further $O(n_{\text{PSF}}^2 n^3)$ calculations in total after each element is calculated in turn. This cost is high: for a single, small, 100×100 input image we have $n = 10^4$, and if we assume that $n_{\text{PSF}} = 3$ we arrive at an estimated $\sim 10^{13}$ floating point operations required for the calculation of $A_{\alpha ij}$.

Another approach is possible. Inspection of equations (17) & (18) shows that we may write expressions for $A_{\alpha ij}$ and $B_{\alpha i}$ directly in terms of inverse Fourier transform operations, as follows:

$$A_{\alpha ij} = \mathcal{F}^{-1} \left[\tilde{G}_i^*(\mathbf{u}) \tilde{G}_j(\mathbf{u}) \right] (\mathbf{r}_j - \mathbf{r}_i), \quad (21)$$

and

$$B_{\alpha i} = -2\mathcal{F}^{-1} \left[\tilde{G}_i(\mathbf{u}) \tilde{\Gamma}^*(\mathbf{u}) \right] (\mathbf{r}_i - \mathbf{R}_\alpha). \quad (22)$$

The determination of $A_{\alpha ij}$ and $B_{\alpha i}$ thus proceeds by first constructing high-resolution lookup tables for these quantities, using the inverse DFT of the $\tilde{G}_i^* \tilde{G}_j$ and $\tilde{G}_i \tilde{\Gamma}^*$ arrays. As additional zero-padding may be added to these complex functions in Fourier space, the lookup tables may be sampled at essentially arbitrary spatial resolution (physical memory constraints notwithstanding) following the inverse DFT back into real space. The value of each element in $A_{\alpha ij}$ and $B_{\alpha i}$ can then be determined by high-order polynomial interpolation (e.g. Press et al. 1992) of the lookup tables to the locations $\mathbf{r}_j - \mathbf{r}_i$ and $\mathbf{r}_i - \mathbf{R}_\alpha$, respectively.

Assuming once more that the PSF images are zero-padded to cover the spatial extent of the input images, the lookup tables will each require $O[n_{\text{PSF}}^2 n \log(n_{\text{PSF}}^2 n)]$ calculations using the FFT. A polynomial interpolation in two dimensions across regularly-gridded data is an operation of complexity $O(2n_{\text{poly}}^2)$, where n_{poly} is the degree of the interpolating polynomial. Therefore, the determination of all elements of $A_{\alpha ij}$ and $B_{\alpha i}$ via lookup table requires $O(2n_{\text{poly}}^2 n^2)$ and $O(2n_{\text{poly}}^2 nm)$ calculations, respectively.⁵ This process will take longer than the time required to construct the lookup tables themselves, but the order n reduction in complexity seen over the direct-summation approach is a significant improvement. It is this technique that is used in the IMCOM prototype; it is noted that both approaches are trivially parallelizable, which is typically the case for calculations made throughout the method.

3.3. Eigendecomposition of the \mathbf{A}_α Matrix

After the calculation of the system matrices of equations (17)–(19), the next computationally challenging task is the solution of equation (20) to find $T_{\alpha i}$. This equation may be rewritten as

$$\mathbf{T}_\alpha (\mathbf{A}_\alpha + \kappa \mathbf{N}) = -\frac{1}{2} \mathbf{B}_\alpha \quad (23)$$

in matrix form; in the case of a single, fixed κ for all α , this system is simply an m -element ensemble of transposed $\mathbf{M}\mathbf{x} = \mathbf{b}$ matrix-vector equations. Here $M_{ij} = A_{\alpha ij} + \kappa N_{ij}$ will be real, symmetric and positive definite, and each solution (corresponding to a row of $T_{\alpha i}$) can be found efficiently and in parallel using the Cholesky decomposition $\mathbf{M} = \mathbf{L}\mathbf{L}^T$ (Press et al. 1992; Anderson et al. 1999). This decomposition need be performed only once for constant κ ; subsequently,

⁵ The authors note that the calculations presented in this paper were in fact made using a polynomial interpolation schema that allows arbitrarily-spaced points along the abscissa, but at a greater relative cost of $O(n_{\text{poly}}^3)$ operations per interpolation.

only the back-substitution step need then be repeated m times, once for each row of $-B_{\alpha i}/2$. The full solution for constant κ therefore requires $O(n^3 + n^2 m)$ calculations.

However, as stated in Section 3.1 the goal of this implementation is to allow the user to specify $\Sigma_{\alpha\alpha}^{\text{max}}$ or U_{α}^{max} for the output image, varying κ from pixel to pixel to satisfy this requirement. If we assume that iteratively finding such a solution requires n_{tries} attempts, then simply solving for $T_{\alpha i}$ using Cholesky decomposition comes at a cost of $O(n_{\text{tries}} n^3 m)$ calculations. We must seek to avoid fourth powers of n and m , as even for small images each of these numbers quickly exceeds $\sim 10^3$.

The problem may also be tackled by considering the eigendecomposition of the system matrix $A_{\alpha ij}$, which is symmetric positive definite and may thus be written as

$$\mathbf{A}_\alpha = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \quad (24)$$

where Q_{ij} is the $n \times n$ matrix whose columns are the n eigenvectors of $A_{\alpha ij}$, and the diagonal matrix $\mathbf{\Lambda}_{ij} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ contains the corresponding eigenvalues (e.g. Arfken & Weber 2005). The inverse is then given simply by

$$\mathbf{A}_\alpha^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^T. \quad (25)$$

This fact can be utilized in the solution for $T_{\alpha ij}$, if we assume that the input noise covariance is diagonal $N_{ij} = \text{diag}(\langle \eta_1^2 \rangle, \langle \eta_2^2 \rangle, \dots, \langle \eta_n^2 \rangle)$. Whether this is a good or poor assumption will depend on the detectors used; off-diagonal terms might occur due to corrections for inter-pixel capacitance (e.g. Barron et al. 2007). It should be noted that even if N_{ij} is not purely diagonal no systematic error is introduced in the output image, since the calculation of the leakage objective U_α is unaffected. The method uses the input noise covariance solely to minimize the output noise, and so imperfect assumptions about N_{ij} will simply cause the resulting $\Sigma_{\alpha\alpha}$ to be less than optimal. Any general N_{ij} can later be propagated through to give the accurate output covariance matrix $\Sigma_{\alpha\beta}$ using equation (9) and the solution for $T_{\alpha i}$ as determined using the diagonal N_{ij} approximation.

Assuming a diagonal input noise N_{ij} , we may then demand that the input image I_i is pre-multiplied by a scaling matrix $\text{diag}(1/\langle \eta_1^2 \rangle^{1/2}, 1/\langle \eta_2^2 \rangle^{1/2}, \dots, 1/\langle \eta_n^2 \rangle^{1/2})$ such that $N_{ij} = \mathbb{I}_n$, the $n \times n$ identity matrix. We then exploit a useful invariance property of the eigenvectors of a symmetric matrix under linear combination with the identity matrix: given equation (24), it is simple to show that

$$\mathbf{A}_\alpha + \kappa_\alpha \mathbb{I}_n = \mathbf{Q} (\mathbf{\Lambda} + \kappa_\alpha \mathbb{I}_n) \mathbf{Q}^T. \quad (26)$$

For our scaled image we may thus write

$$(\mathbf{A}_\alpha + \kappa_\alpha \mathbf{N})^{-1} = \mathbf{Q} (\mathbf{\Lambda} + \kappa_\alpha \mathbb{I}_n)^{-1} \mathbf{Q}^T, \quad (27)$$

where the central matrix is purely diagonal and trivially calculated.

The number of operations required for the initial decomposition of equation (24) is $O(n^3)$, and this decomposition need only be performed once and may be parallelized (e.g. Anderson et al. 1999). We show in the following Section 3.4 that, subsequently, trial values of $\Sigma_{\alpha\alpha}(\kappa_\alpha)$ or $U_\alpha(\kappa_\alpha)$ may be calculated for each output pixel α at a cost of $O(n)$ operations per trial solution. This implies a total cost of $O(n^3 + 2n^2 m + n_{\text{tries}} nm)$ for calculating $T_{\alpha ij}$, including an extra $O(n^2 m)$ overhead for necessary matrix manipulation as

will be discussed in the following Section. This represents a significant economy over the repeated use of Cholesky decompositions and justifies the use of the eigendecomposition. This procedure also yields a figure for the condition number for the matrix $A_{\alpha ij}$, which gives a useful indication of the accuracy and stability of results regarding the input system.

3.4. Solution for κ by Interval Bisection

Having calculated the eigenvectors Q_{ij} and eigenvalues Λ_i of the system matrix $A_{\alpha ij}$, we see from equation (27) that these may be used to trivially express the matrix inverse required to calculate $T_{\alpha i}$ in equation (20). In Appendix A, equations (A3) & (A4) also give expressions for $\Sigma_{\alpha\alpha}$ and U_α in terms of this matrix inverse. Substituting equation (27) and expanding, we find

$$\Sigma_{\alpha\alpha} = \frac{1}{4} \sum_i \frac{N_{ii}}{(\lambda_i + \kappa_\alpha)^2} \left(\sum_j B_{\alpha j} Q_{ji} \right)^2, \quad (28)$$

and

$$U_\alpha = C_\alpha - \frac{1}{4} \sum_i \frac{\lambda_i + 2\kappa_\alpha}{(\lambda_i + \kappa_\alpha)^2} \left(\sum_j B_{\alpha j} Q_{ji} \right)^2, \quad (29)$$

where here we have chosen to write the matrix multiplication explicitly for clarity. These expressions may be used to generate trial values of $\Sigma_{\alpha\alpha}(\kappa_\alpha)$ or $U_\alpha(\kappa_\alpha)$ directly, without first calculating $T_{\alpha ij}$. The object

$$P_{\alpha i} = \sum_j B_{\alpha j} Q_{ji} \quad (30)$$

will be referred to as the projection matrix, and gives the scalar product of each row-vector \mathbf{B}_α with the successive eigenvectors Q_{ij} of $A_{\alpha ij}$. This projection matrix can be calculated in advance, immediately following the eigendecomposition of $A_{\alpha ij}$, at a cost of $O(n^2 m)$ operations. Once this is done, each trial calculation of $\Sigma_{\alpha\alpha}$ or U_α made using equations (28) & (29) comes at the low cost of $O(n)$ operations. After each κ_α is found, we may then use the projection matrix once more to write the final solution for $T_{\alpha i}$ as

$$T_{\alpha i} = -\frac{1}{2} \sum_j \frac{P_{\alpha j} Q_{ij}}{\lambda_j + \kappa_\alpha}, \quad (31)$$

at a further cost of $O(n^2 m)$ operations.

The problem is then, for each output pixel α , to find the value of κ_α that satisfies $\Sigma_{\alpha\alpha}^{\max} - \Delta \Sigma_{\alpha\alpha}^{\max} < \Sigma_{\alpha\alpha} \leq \Sigma_{\alpha\alpha}^{\max}$ or $U_\alpha^{\max} - \Delta U_\alpha^{\max} < U_\alpha \leq U_\alpha^{\max}$. The simplest and most robust algorithm for finding such a solution is interval bisection (Press et al. 1992), and this is the method we adopt here, iteratively bisecting some search range $[\kappa_{\min}, \kappa_{\max}]$ at intervals of $\log \kappa$. If a root lies within this range the algorithm will find it, or will get as close as allowed by the user-input maximum number of bisections n_{bis} (see Section 3.1, and Section 4.3 for some example choices of these parameters).

Since there is some known directionality in the problem at hand, the algorithm may be adapted somewhat. In Appendix A, we show that the derivatives of $\Sigma_{\alpha\alpha}$ or U_α with respect to κ are everywhere negative or positive, respectively. Taking the case of U_α , this means that the first trial value should be calculated using κ_{\min} : if this solution does not satisfy $U_\alpha \leq U_\alpha^{\max}$ then there can be no improvement on κ_{\min}

and this value is adopted as κ_α . For $\Sigma_{\alpha\alpha}$, we likewise start on κ_{\max} before proceeding to bisection. An output image of the interval bisection-derived κ_α is also generated, so the user may see if this range needs to be expanded. In addition, it is also relatively fast to generate an image of the U_α derivative given in equation (A6), which allows the user to judge convergence upon the fundamentally limiting lower bound, $\lim_{\kappa \rightarrow 0^+} U_\alpha$.

The total cost of solving for $T_{\alpha ij}$ is therefore $O(n^3 + 2n^2 m + n_{\text{tries}} n m)$ as given in Section 3.3, which includes the eigendecomposition of $A_{\alpha ij}$, the calculation of the projection matrix, the n_{tries} trial solutions to κ_α for each of m output pixels, and the final determination of $T_{\alpha ij}$ using equation (31). It can be seen that the dominant contributions come from the eigendecomposition of $A_{\alpha ij}$ and calculating $P_{\alpha i}$, which lessens the need to seek more efficient (and potentially less robust) algorithms for the subsequent root-finding.

4. AN ILLUSTRATIVE EXAMPLE

We now demonstrate the use of the method by means of a worked example. As a primary goal of this work is the analysis of survey images for a dark energy mission, we perform a simple simulation of observations made using an example design concept under consideration for the *WFIRST* mission. This particular design employs a $D = 1.3$ m primary mirror with an off-axis configuration for the secondary mirrors and detector array. This provides an unobstructed aperture, both increasing the flux throughput of the instrument and reducing the integrated light in the outer wings of the PSF.⁶

We will simulate monochromatic observations at the near-infrared wavelength $\lambda = 1 \mu\text{m}$, close to the center of the range proposed for the *WFIRST* imaging survey (Gehrels 2010). It should be noted that observations may likely take place at longer wavelengths than this value, and over a broad filter band. This will have two desirable effects, smearing together diffraction rings in the PSF and introducing a cutoff at lower spatial frequencies in the MTF. The $\lambda = 1 \mu\text{m}$ monochromatic system adopted for these examples is, in this sense, more demanding than we expect for a realistic dark energy mission.

In Figure 1 (left panel) we show a simulated PSF, generated using the ZEMAX software (Ed Cheng, John Lehan, David Content, & the *WFIRST* Project Office, priv. comm.), for this *WFIRST* design concept at this wavelength. The model includes additional astigmatism to mimic aberrations due to fabrication errors, and coma to simulate possible misalignments due to thermal drift and realistic imperfections in installation and testing. For the camera we assume an array of $18 \mu\text{m}$ Teledyne Hawaii-2RG (H2RG) detectors, corresponding to an image sampling of 0.18 arcsec. For the purposes of this demonstration we model the pixel response as a simple boxcar function.

We also model the effects of charge diffusion using results from Barron et al. (2007), who measure a projected, one-dimensional diffusion length of $l_{\text{CD}} = 1.87 \mu\text{m}$ for a hyperbolic secant diffusion function $I_{\text{CD}}(\Delta x) \propto \text{sech}(\Delta x / l_{\text{CD}})$, where Δx is the distance of the collected charge from the location of the electron-hole pair. Since we seek a de-projected charge diffusion model, some assumptions are unavoidable: for simplicity, we model the two-dimensional charge diffusion as a zero mean, circular Gaussian of width $\sigma_{\text{CD}} = \pi l_{\text{CD}} / 2$, preserving the variance (and hence diffusion length)

⁶ More details regarding this family of designs can be found in the report of the Interim Science Working Group, http://wfirst.gsfc.nasa.gov/science/ISWG_Report.pdf

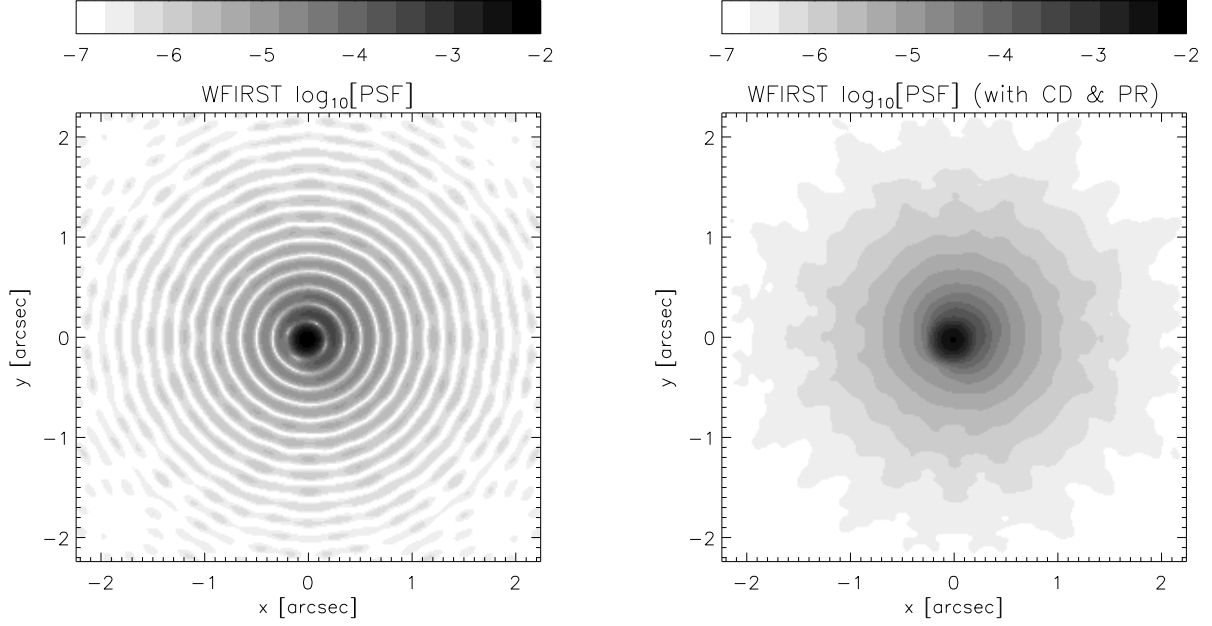


Figure 1. Left panel: Simulated PSF for a *WFIRST* mission concept with a 1.3 m unobstructed primary at a wavelength of $1\ \mu\text{m}$. Right panel: Image of $G_i(\mathbf{r})$ for this PSF, incorporating additional charge diffusion (CD) and the pixel response (PR) for 0.18 arcsec detectors. Both images use a logarithmic scale.

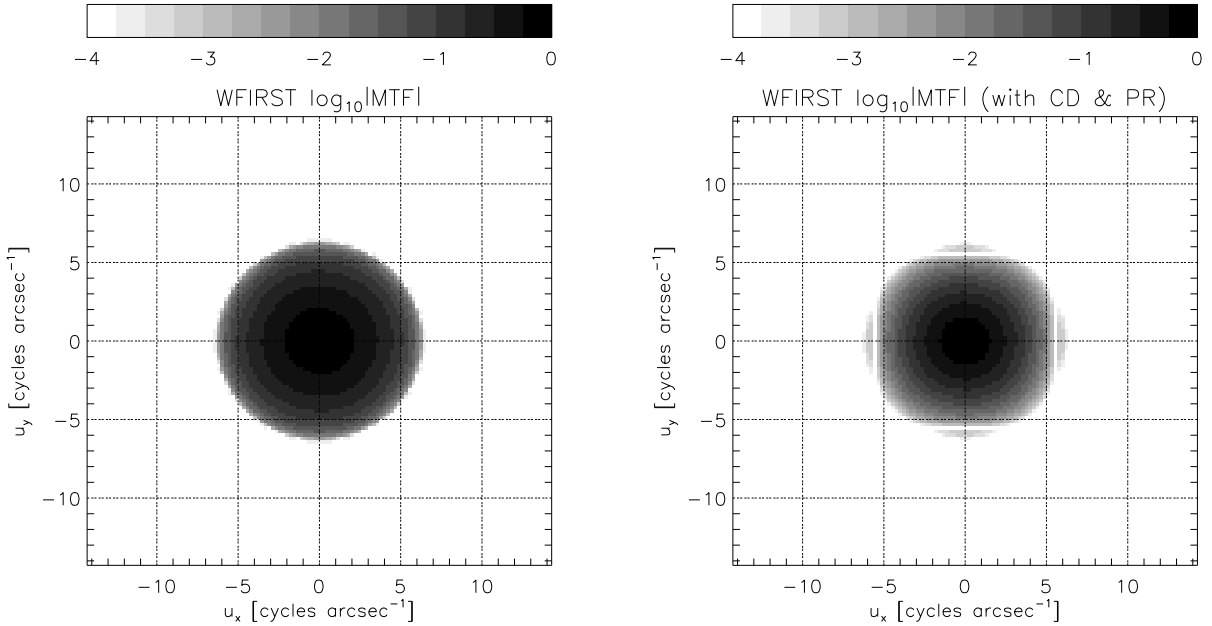


Figure 2. Left panel: Magnitude of the Modulation Transfer Function (MTF) for the *WFIRST* concept PSF of Figure 1. Right panel: Image of $|\tilde{G}_i(\mathbf{u})|$, the magnitude of the MTF corresponding to $G_i(\mathbf{r})$; the sinc modulation of the pixel response can be clearly seen at $1/0.18 = 5\frac{5}{9}$ cycles arcsec $^{-1}$. Both images use a logarithmic scale.

of the function.

Convolving the *WFIRST* concept PSF with the pixel response boxcar and the charge diffusion function, we arrive at the functional form of $G_i(\mathbf{r})$ for this demonstration, shown in Figure 1 (right panel). For this first demonstration, $G_i(\mathbf{r})$ will remain constant throughout the input image.

The Modulation Transfer Function (MTF) is defined as the Fourier transform of the telescope PSF. The function $\tilde{G}_i(\mathbf{u})$ therefore represents the MTF conjugate to $G_i(\mathbf{r})$, and the

magnitude of this complex object can be seen in Figure 2 for the two corresponding PSFs of Figure 1. As can be seen from Figure 2, the system is bandlimited at the fundamental frequency corresponding to $D/\lambda = 6.3026$ cycles arcsec $^{-1}$. Therefore, a sampling interval of $\lambda/2D = 0.079333$ arcsec in the output image H_α is the requirement for critical sampling according to the sampling theorem (Marks 2009). This output sampling rate is therefore of fundamental interest in demonstrating that survey design and multiple dithering strategies

will allow a dark energy mission to overcome an undersampling detector array in the focal plane, and thereby utilize all available spatial information for weak lensing measurements. Hence, $\lambda/2D = 0.079333$ arcsec will be adopted as the desired output sampling interval for all the demonstrations in this study.

4.1. Input Images: Random Dithers, Random Rotations

From the *WFIRST* concept described in Section 4 it has been possible to create a model for $G_i(\mathbf{r})$ that shares many of the properties expected for a future dark energy mission. However, this represents only one of the input quantities described in Section 3.1; not least, input images themselves are required.

As a demonstration of the technique's flexibility we will attempt to reconstruct a fully-sampled output image using a set of three, randomly dithered and rotated images. The spatial configuration \mathbf{r}_i of the three dithered exposures can be seen in Figure 3, each of dimension 20×20 pixels². Also shown are the corresponding sample locations \mathbf{R}_α of the desired, fully-sampled output image. It should be stated from the outset that these three dither patterns are not expected to be sufficient to achieve full sampling, but are being used deliberately to show the behavior of the algorithm when unable to recover certain output pixels. The user may of course use images with any pixel locations, there being no requirements of regularity or homogeneity in the formalism.

For the physical image on the sky $f(\mathbf{r})$ we choose a single galaxy, modeled as a simple elliptical, exponential profile. The profile is given an axis ratio $a/b = 3/\sqrt{5}$, a semi-major axis scale length (aligned along the x direction) of $a = 3/(4\sqrt{5})$ arcsec, and a central peak surface brightness of 50 (in arbitrary units). This model galaxy is shown in Figure 4 in high-resolution.

This image is then convolved with the $G_i(\mathbf{r})$ of Figure 1 and sampled at the locations shown in Figure 3 to generate the array of input images that make up I_i . Since each image is rotated, before calculating $A_{\alpha ij}$ and $B_{\alpha i}$ the PSF $G_i(\mathbf{r})$ for each exposure must be rotated into a common coordinate system in which $\Gamma(\mathbf{r})$ is defined (see Section 4.2 below). In this example we choose $\Gamma(\mathbf{r})$ to be simply the unrotated $G_i(\mathbf{r})$ of Figure 1, so that only the input PSFs need be rotated before solving the system. For precise comparison of output and input images we do not add stochastic noise to I_i , but will nonetheless assume a unit diagonal noise covariance $N_{ij} = \mathbb{I}_n$ to illustrate the behavior of $\Sigma_{\alpha\alpha}$ in a more realistic, noisy imaging scenario.

With $G_i(\mathbf{r})$, I_i , \mathbf{r}_i and N_{ij} we have provided the three “hard inputs” of Section 3.1, but there remain choices to be made regarding the values of user-specified input quantities. We adopt a simple Dirac delta function for $\Upsilon(\mathbf{r})$ so that $\tilde{\Upsilon}(\mathbf{u}) = 1$. This fixed parameter choice is given in Table 1, where all such implementation-specific parameter values are given. The remaining choices that must be made relate to the process of finding solutions for U_α^{\max} or Σ_α^{\max} , and so we defer a discussion of these input quantities until Section 4.3. We first illustrate the system matrices $A_{\alpha ij}$, $B_{\alpha i}$, which may be calculated immediately.

4.2. System Matrices

Taking the input functional forms of $G_i(\mathbf{r})$ and $\Gamma(\mathbf{r})$, and the pixel locations \mathbf{r}_i and \mathbf{R}_α , we may calculate $A_{\alpha ij}$ and $B_{\alpha i}$ using the FFT-based prescription described in Section 3.2. As

the accuracy of these objects is crucial, and their computation costs are relatively low in comparison to later operations, it is desirable to generate high-resolution lookup tables for $A_{\alpha ij}$ and $B_{\alpha i}$ and then interpolate using a high-order polynomial.

As discussed in Section 3.2 the images of $G_i(\mathbf{r})$ and $\Gamma(\mathbf{r})$ are first zero-padded so that their spatial dimensions (in physical units) equal or exceed the maximum spatial extent of the input images that make up I_i . It is noted that the *WFIRST* design concept PSF of Figure 1 is supplied at relatively high resolution ($n_{\text{PSF}} \simeq 11$: see Section 3.2). For rotated input frames, each $G_i(\mathbf{r})$ must be rotated accordingly to the common coordinate system in which $\Gamma(\mathbf{r})$ is defined. Each image $G_i(\mathbf{r})$ is rotated using Lagrange polynomial interpolation at order $n_{\text{poly}} = 7$. Then $\tilde{G}_i(\mathbf{u})$ and $\tilde{\Gamma}(\mathbf{r})$ are calculated using the FFT algorithm. Limited only by the memory available (2GB in a 32-bit architecture), we may then add *further* zero-padding to the bandlimited $\tilde{G}_i(\mathbf{u})$ and $\tilde{\Gamma}(\mathbf{r})$. We choose to zero-pad each such that their linear dimension increases by a further factor n_{pad} compared to the starting size of $\tilde{G}_i(\mathbf{u})$ and $\tilde{\Gamma}(\mathbf{r})$ (which were already padded once in real space, up to the size of the input images). In this example, and throughout this Paper, we choose $n_{\text{pad}} = 3$ (see Table 1).

The inverse FFT is then used to build the lookup tables described by equations (21) & (22) which, following the zero-padding, have $n_{\text{PSF}} \times n_{\text{pad}}$ entries along the length of each input pixel in each dimension. Finally, these tables are interpolated using a polynomial at order $n_{\text{poly}} = 7$ to estimate values of $A_{\alpha ij}$ and $B_{\alpha i}$ at the locations of the vectors $\mathbf{r}_j - \mathbf{r}_i$ and $\mathbf{r}_i - \mathbf{R}_\alpha$, respectively. The resultant matrices can be seen in Figures 5 & 6.

Once $A_{\alpha ij}$ and $B_{\alpha i}$ are calculated, the final stages in the preparation for the solution of $T_{\alpha i}$ are the eigendecomposition of the $A_{\alpha ij}$ matrix to give λ_i and Q_{ij} , and the subsequent calculation of the projection matrix $P_{\alpha i}$. In Figures 7 & 8 we plot the eigenvalues and lowest-order eigenvectors for $A_{\alpha ij}$ in this example. The condition number \aleph (see, e.g., Cheney & Kincaid 1998) for a symmetric, positive-definite matrix such as $A_{\alpha ij}$ is simply expressed in terms of the eigenvalues as

$$\aleph(A_{\alpha ij}) = \frac{\max[\lambda_i]}{\min[\lambda_i]}. \quad (32)$$

A system is said to be singular if the condition number is infinite, and typically ill-conditioned if $\log_{10}(\aleph) \gtrsim$ the numerical precision of at which each matrix element is stored. For the illustrative example in this Section, the condition number is found to be $\aleph(A_{\alpha ij}) = 8.369 \times 10^6$, which implies the system is adequately well-conditioned if using double precision arithmetic with machine epsilon $\varepsilon = 2^{-53} = 1.11 \times 10^{-16}$. In practice the matrix that must be inverted is $A_{\alpha ij} + \kappa_\alpha N_{ij}$ so, given appropriate choices of κ_α , the method may proceed even when \aleph is large. Nonetheless, calculating \aleph in this way provides both a useful check on the eigenvalues λ_i (which must all be positive since $A_{\alpha ij}$ is a symmetric positive definite matrix), and an idea of when solutions for $T_{\alpha i}$ might become unstable for small κ_α .

The projection matrix $P_{\alpha i}$ is then constructed from the eigenvectors Q_{ij} and the matrix $B_{\alpha i}$ via matrix multiplication as shown in equation (30). At this stage the problem is then fully set up, and we are ready to begin the solution for $T_{\alpha i}$.

4.3. Output Images

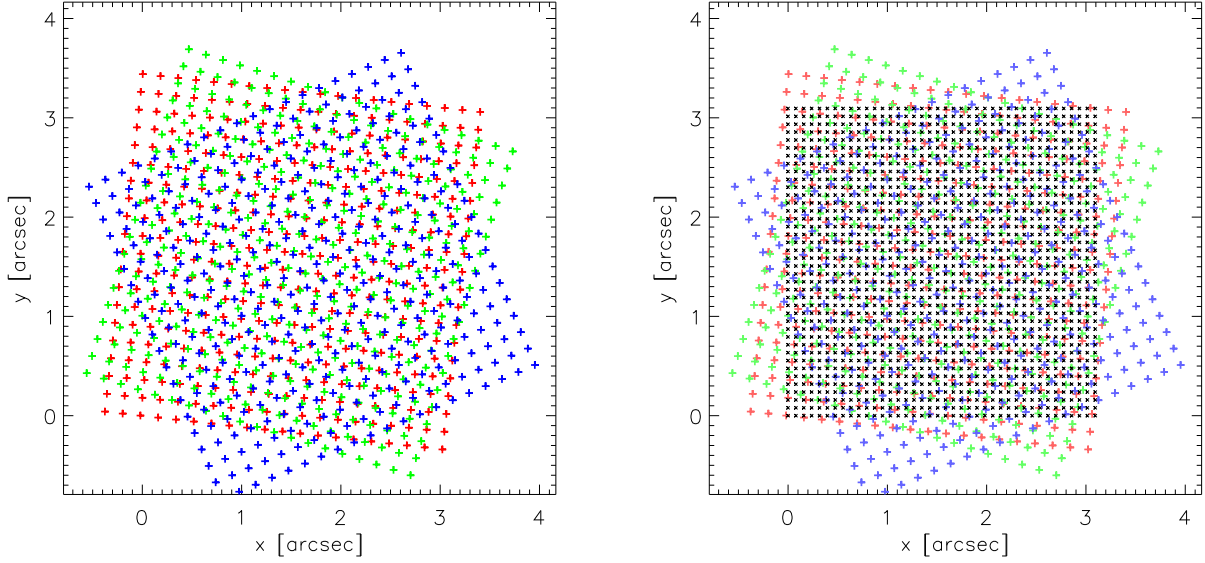


Figure 3. Left panel: Input pixel locations \mathbf{r}_i for three randomly dithered and rotated exposures of the same camera. Red, green and blue crosses represent each of the three exposures. Right panel: Output image sampling locations \mathbf{R}_α (black crosses), placed at a spatial interval of 0.079333 arcsec so as to fully sample the *WFIRST* concept image plane.

Table 1

Fixed, IMCOM implementation-specific parameters used in calculating the results of Sections 4 & 5.

Parameter	Value	Description
$\Upsilon(\mathbf{r})$	$\delta(\mathbf{r})$	Kernel used for calculating leakage objective U_α (see equation 11).
n_{pad}	3	Multiple by which $\tilde{G}_i(\mathbf{u})$ and $\tilde{\Gamma}(\mathbf{u})$ are <i>further</i> zero-padded in \mathbf{u} when calculating $A_{\alpha ij}$ and $B_{\alpha i}$ (Section 4.2).
n_{poly}	7	Polynomial order used when rotating $G_i(\mathbf{r})$, and interpolating $A_{\alpha ij}$ and $B_{\alpha i}$ from lookup tables (Section 4.2).
κ_{min}	$1.11C_\alpha \times 10^{-16}$	Minimum κ_α for interval bisection search, set by machine precision (Sections 3 & 4.3).
κ_{max}	$9.01C_\alpha \times 10^{15}$	Maximum κ_α for interval bisection search, set by machine precision (Sections 3 & 4.3).
n_{bis}	53	Maximum number of interval bisections allowed in solving for each κ_α (Sections 3 & 4.3).

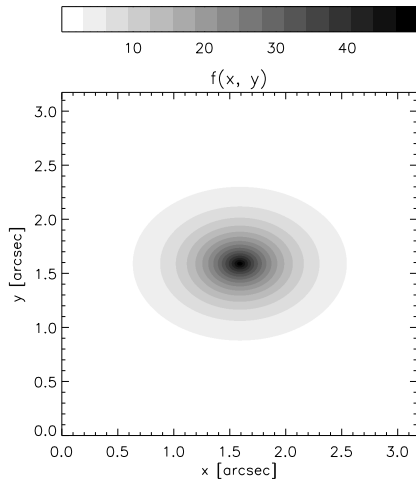


Figure 4. Surface brightness $f(\mathbf{r})$ used in Sections 4 & 5: a simple, idealized model of an elliptical galaxy.

The solution of equation (23) for $T_{\alpha i}$ as described in Section 3.4 proceeds in two separate stages. First, values of κ_α

for each output pixel are independently determined so as to attempt to keep U_α or $\Sigma_{\alpha\alpha}$ within the pre-set limits. Second, and at far greater relative computational cost, these values are substituted into equation (31) to calculate $T_{\alpha i}$.

To determine κ_α , it is necessary to complete the set of choices described in Section 3.1. For the purposes of this example, we will demonstrate the solution of κ_α both by setting requirements upon $\Sigma_{\alpha\alpha}$ and by setting requirements on U_α . For both cases, we search within the broad range $[\kappa_\alpha^{\text{min}}, \kappa_\alpha^{\text{max}}] = [1.11C_\alpha \times 10^{-16}, 9.01C_\alpha \times 10^{15}]$ (see Table 1), effectively being limited only by the machine precision ε . We also allow a total of $n_{\text{bis}} = 53$ interval bisections to find κ_α . Computational costs are negligibly affected by choices for these parameters.

In Figure 9 we plot results from the IMCOM code having required $\Sigma_{\alpha\alpha}^{\text{max}} - \Delta\Sigma_{\alpha\alpha}^{\text{max}} \leq \Sigma_{\alpha\alpha} < \Sigma_{\alpha\alpha}^{\text{max}}$, where $\Sigma_{\alpha\alpha}^{\text{max}} = 1$ and $\Delta\Sigma_{\alpha\alpha}^{\text{max}} = 10^{-2}$. As the input noise covariance is in this case specified as a unitary, diagonal $N_{ij} = \mathbb{I}_n$, this corresponds to requiring that the output image H_α maintain the same level of noise variance as compared to the input image pixels. Figure 9a shows the value of κ_α selected at each output position \mathbf{R}_α in order to fulfill these requirements on $\Sigma_{\alpha\alpha}$, and Figure

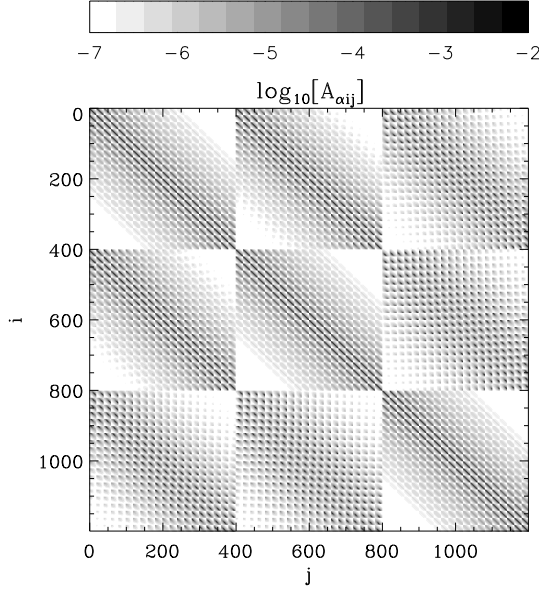


Figure 5. System matrix $A_{\alpha ij}$ for the image configuration of Section 4 and Figure 10.

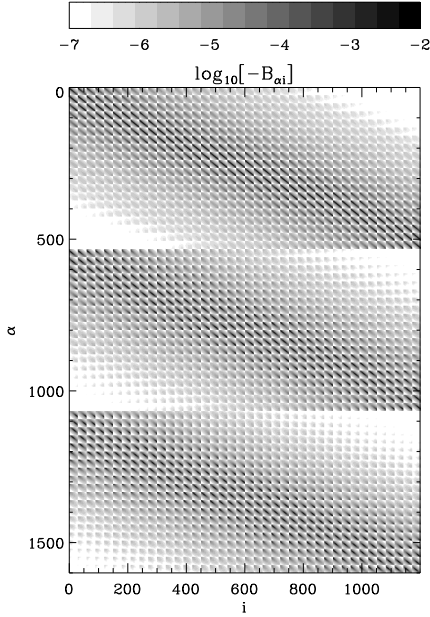


Figure 6. System matrix $B_{\alpha i}$ for the image configuration of Section 4 and Figure 10.

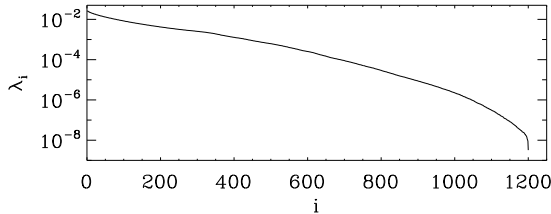


Figure 7. Eigenvalues of the $A_{\alpha ij}$ matrix shown in Figure 5. The matrix in this example has a condition number of $\aleph = 8.369 \times 10^6$.

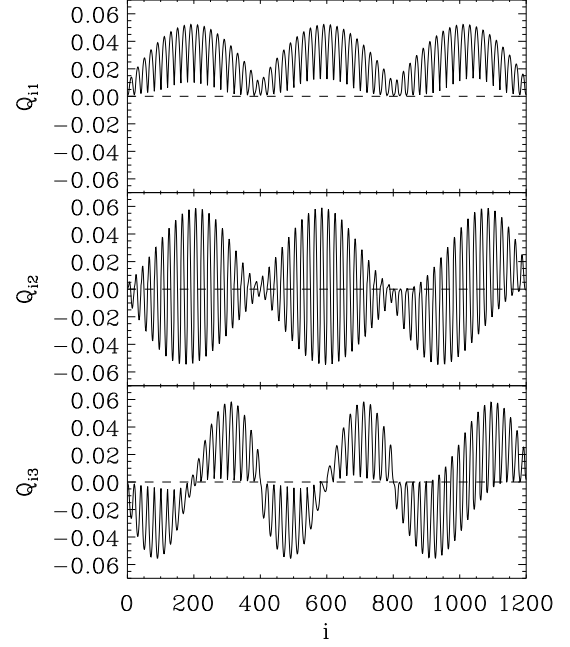


Figure 8. The first three eigenvectors of the $A_{\alpha ij}$ matrix shown in Figure 5.

9b shows the variation in the resultant leakage objective U_α across the output image. This latter is plotted in its natural units of C_α (see Appendix A, equations A8 & A9). Degradation in image fidelity, manifesting as increased U_α , can be clearly seen in the corner regions of the output in Figure 9b where input pixel coverage is reduced, but also within central locations where chance alignments of the input pixel pattern lead to encircled regions of sparse sampling.

Overall we see that, given the requirements on the noise properties, the randomly dithered, three exposure system produces output with $U_\alpha \sim 0.01C_\alpha$ or worse over a significant fraction of H_α . This is not the level of fidelity that will be required for a dark energy mission, but the demonstration illustrates how aspects of the input dither pattern impact the quality of the recovered output.

The corresponding noise variance map $\Sigma_{\alpha\alpha}$ can be seen in Figure 9c. As shown in Appendix A, the output $\Sigma_{\alpha\alpha}$ can be made arbitrarily small by increasing κ_α , but not necessarily arbitrarily large as κ_α decreases. The map of $\Sigma_{\alpha\alpha}$ shows where occasional chance alignments of multiple input pixels produce noise that is *lower* than $\Sigma_{\alpha\alpha}^{\max} - \Delta\Sigma_{\alpha\alpha}^{\max}$, and comparison with Figure 9a shows that these points indeed lie where $\kappa_\alpha = \kappa_\alpha^{\min}$.

In Figure 9d we compare the IMCOM-reconstructed image H_α to an independently-generated target image J_α , calculated using equation (3) and the highly-oversampled $f(\mathbf{r})$ as shown in Figure 4 (the code for generating J_α is in fact written in the IDL language and shares no routines with IMCOM). These two image vectors are used to calculate the squared fractional residual, which we define as

$$F_\alpha^2 = \frac{(H_\alpha - J_\alpha)^2}{J_\alpha^2}. \quad (33)$$

Values of F_α^2 depend upon $f(\mathbf{r})$ directly and so cannot themselves be used to locate an optimal $T_{\alpha i}$, but in simulated examples they provide a useful comparison: for a well-controlled system we typically expect $F_\alpha^2 \lesssim U_\alpha/C_\alpha$ [al-

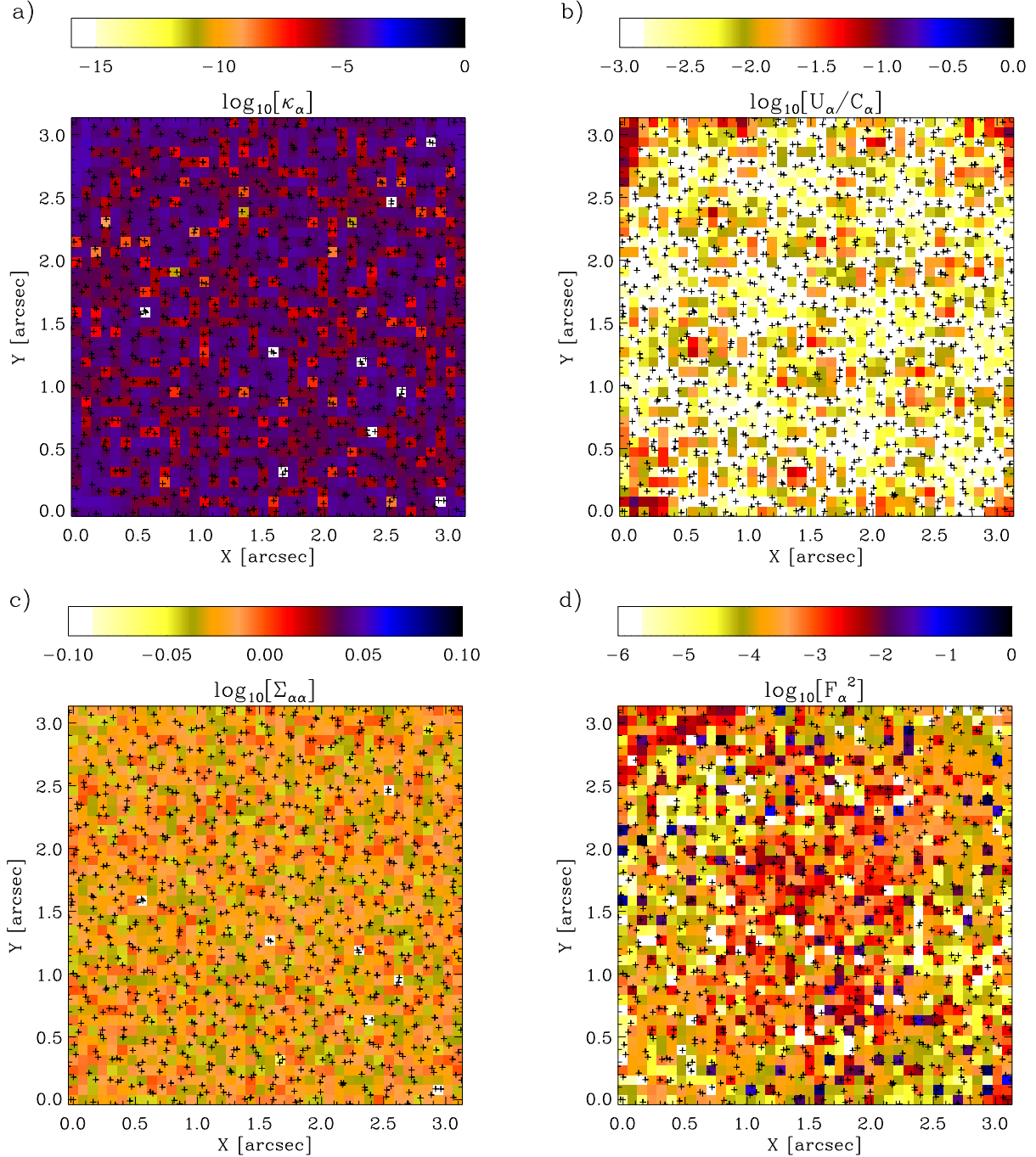


Figure 9. Map of κ_α solution (a), output leakage objective U_α (b), noise variance $\Sigma_{\alpha\alpha}$ (c), and corresponding F_α^2 (d) for the input choices $\Sigma_{\alpha\alpha}^{\max} = 10^{-1}$, $\Delta\Sigma_{\alpha\alpha}^{\max} = 10^{-3}$ in the demonstration of Section 4. The color scale in the map for U_α does not reflect the full range of these values, but is selected to most clearly illustrate the dependence of U_α on the input dither pattern \mathbf{r}_i .

though this may be violated in practice if $J_\alpha \simeq 0$ or $f(\mathbf{r})$ contains large or abrupt intensity variations such as those caused by bright stars or cosmic ray impacts]. The variation in F_α^2 across the output image plane for this example can be seen in Figure 9d. This shows that $F_\alpha^2 \lesssim U_\alpha/C_\alpha$ across much of the output in this example but can vary sharply due to the large variations in U_α , particularly towards the edge regions of the image where $f(\mathbf{r})$ is smaller.

We now examine a case in which the desired output im-

age properties are specified in terms of the leakage objective U_α rather than the noise variance. In Figure 10 we plot results from the IMCOM code having required $U_\alpha^{\max} - \Delta U_\alpha^{\max} \leq U_\alpha < U_\alpha^{\max}$, where $U_\alpha^{\max} = 10^{-2}C_\alpha$ and $\Delta U_\alpha^{\max} = 10^{-4}C_\alpha$. It can be seen from Figure 10b that in fact $U_\alpha > 10^{-2}C_\alpha$ in the sparsely-sampled corner regions of the output region. Here in the corners the system has saturated at the lower limit for U_α within the broad range $[\kappa_\alpha^{\min}, \kappa_\alpha^{\max}]$, the practical minimum possible using double precision arithmetic and possibly

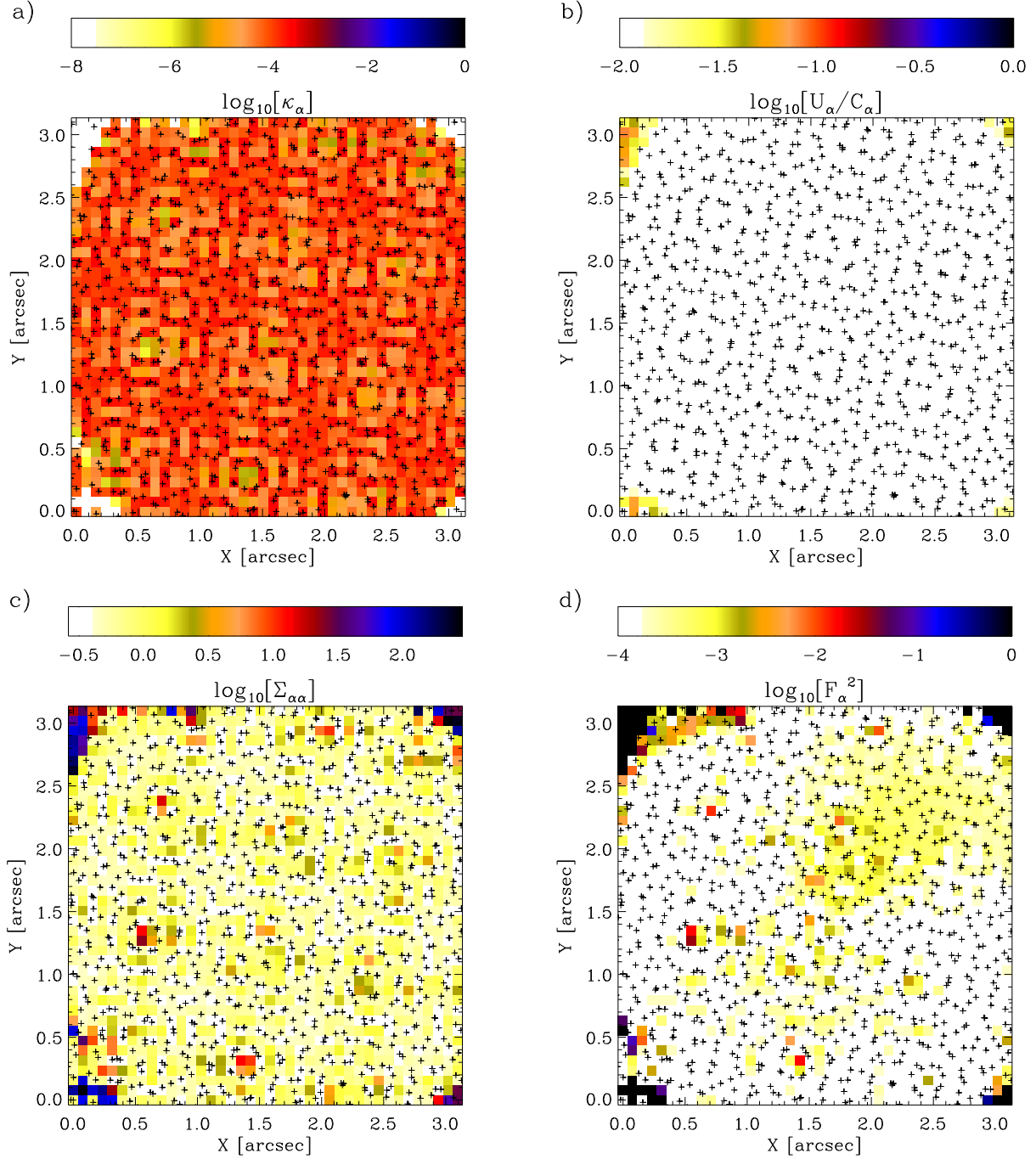


Figure 10. Map of κ_α solution (a), output leakage objective U_α (b), noise variance $\Sigma_{\alpha\alpha}$ (c), and corresponding F_α^2 (d) for the input choices $U_\alpha^{\max} = 10^{-2}$, $\Delta U_\alpha^{\max} = 10^{-4}$ in the demonstration of Section 4. The failure to reach $U_\alpha \leq 1 \times 10^{-2} C_\alpha$ in corner regions, even for $\kappa_\alpha = \kappa_\alpha^{\min}$, can be clearly seen; the localization of peaks in the noise variance around encircled zones of reduced pixel coverage is also apparent.

approaching the theoretical minimum given by equation (A9).

Figure 10c shows the output $\Sigma_{\alpha\alpha}$ for this solution: rapid variation is seen in the noise properties for this output image, with factors of ~ 10 – 100 greater variance than the input \tilde{N}_{ij} seen in some regions. These include the corner regions but also, as was the case for U_α in the previous example, encircled regions of sparse sampling that are reflected in the output $\Sigma_{\alpha\alpha}$. There are also similar traces of these effects in the F_α^2 map of Figure 10d, although these are most marked in the

edge and corner regions where sparse input sampling combines with small values of J_α in the target image.

Overall, it is seen that restricting $U_\alpha < 10^{-2} C_\alpha$ (a rather modest ambition if attempting precision scientific work) is difficult for the input image configuration and PSF of this example system. While, as shown in Figure 9, it was possible to create an image with benign noise properties, this comes at a cost of significant U_α across much of the image. Large values for U_α require significant leakage $L_\alpha(\mathbf{R}_\alpha - \mathbf{r})$, and therefore

imply a non-negligible, biasing residual $\langle Z_\alpha \rangle$. The input sampling is too sparse to allow stable, unbiased reconstruction at this output resolution. While this would have been apparent from the outset to someone with experience in stacking imaging data, the linear formalism allows it to be quantitatively diagnosed using $\Sigma_{\alpha\alpha}$ and U_α in a way that is independent of $f(\mathbf{r})$. This capability makes the method useful as a survey design tool.

In this Section we have attempted to demonstrate the flexibility and power of the linear image combination formalism developed in Section 2 by tackling a difficult case. **For the three randomly-dithered exposure example, where the input sampling turns out to be insufficient for a fully-sampled, unbiased output, we have illustrated how the objective quantities U_α and $\Sigma_{\alpha\alpha}$ are used to understand the properties of H_α .** In placing simple, user-specified constraints on these objective quantities we have shown how the IMCOM implementation of Section 3 can be used to efficiently explore the optimal trade-off between noise and fidelity, using a freely-varying κ_α . In trickier cases such as that presented here, there are further synthetic options that might be used to improve noise and bias in H_α : specifying $\Gamma(\mathbf{r})$ as a broader PSF is one possibility (see also Section 5.5).

For the purposes of designing a dark energy mission, however, it is useful to address the problem in a slightly different manner. Given a telescope PSF and an undersampling focal plane detector configuration, we may ask how many dithered exposures are needed, and in which pattern, to generate an unbiased, fully-sampled output H_α . This represents an important practical application of the techniques developed in Sections 2 & 3, and it is using this capacity as a design tool that we explore a range of dither patterns in the following Section.

5. TESTS OF MULTIPLE OBSERVING SCENARIOS

In this Section we assess the ability of a number of dither pattern configurations to reconstruct fully-sampled images with a low level of leakage, a sure requirement for the precision photometry and shape measurement that will be necessary for the imaging component of a dark energy mission. For the purposes of these tests we define low leakage as having $U_\alpha < 10^{-8}C_\alpha$, implying control of the output PSF properties to one part in 10^4 . This ensures that uncontrolled changes to the PSF from linear image reconstruction will account for only a fraction of the total systematic error budget for a dark energy survey (Amara & Réfrégier 2008). We then examine the capability of the *WFIRST* design concept described in Section 4, with an optical PSF given by Figure 1 and a focal plane of H2RG detectors sampling at 0.18 arcsec, to generate fully-sampled outputs that are unbiased at this level of leakage.

The problem is set up as follows: We adopt $U_\alpha^{\max} = 10^{-8}C_\alpha$ and $\Delta U_\alpha^{\max} = 10^{-10}C_\alpha$ as our requirements on the properties of the output image H_α , whereas other inputs such as the undersampling input pixel scale (0.18 arcsec), $G_i(\mathbf{r})$, $\Gamma(\mathbf{r})$, and the parameters of Table 1, are all as described in Section 4 *unless explicitly stated as otherwise*. Using a realistic *WFIRST* design concept in this way, we can begin to explore what survey design strategies may be needed to overcome undersampling for a dark energy mission of this sort. We continue to set $N_{ij} = \mathbb{I}_n$ for the easy interpretation of $\Sigma_{\alpha\alpha}$ results.

The output \mathbf{R}_α will again be placed at an interval of 0.079333 arcsec, the requirement for critical sampling. In

each following test we then vary the input pixel sampling information, adopting different dither patterns, numbers of exposures or detector coverage: in terms of the formalism, these variations are manifested as changes in \mathbf{r}_i and I_i . In this way we seek to show how the linear image combination formalism can be used to inform survey design, and make recommendations about survey dithering strategies for this *WFIRST* design concept.

5.1. The 2×2 Ideal Dither

In Figure 11 we plot results for four 36×36 pixel² input exposures, configured in a 2×2 ideal dither pattern. The upper-left panel, Figure 11a, illustrates this dithered configuration of input exposures (colored points) along with output sampling locations (black crosses), focused in upon a small sub-section of the total image area for clarity. Figure 11b shows the minimum leakage objective U_α that may be achieved at fully-sampled resolution for this dither pattern; as can be seen, the threshold value of $U_\alpha^{\max} = 10^{-8}C_\alpha$ is often exceeded (it is reached only where output pixels closely align with input pixel locations, indicated by the regularly spaced minima in U_α).

The map in Figure 11c shows the output noise variance $\Sigma_{\alpha\alpha}$ for the 2×2 dithered image. Interestingly, this is both reasonably stationary (i.e. approximately constant), and comparable in amplitude to the input noise variance, across a large area of the output. We note that in this Figure, and in many subsequent plots of $\Sigma_{\alpha\alpha}$, we choose the color scale not to show the full dynamic range of $\Sigma_{\alpha\alpha}$ but to better highlight behavior in the central regions of the image.

At the extreme edges of the output, where input information is most scarce, $\Sigma_{\alpha\alpha}$ may exceed 10^5 . The size of such noisy edge regions is primarily determined by the extent of the PSF $G_i(\mathbf{r})$ within each contributing exposure. However, as will be discussed in Section 6, these regions are not likely to be used in a realistic survey strategy: Figure 18 illustrates that it is not necessary for these edge regions to fall within the output H_α at all. Instead, by tessellating successive H_α that cover only the central regions of overlapping input patches, a contiguous output image can be constructed for which all pixels are sufficiently far from an input edge. In this Section, however, these noisier edge regions are kept so as to demonstrate the effect.

Figure 11d shows the map of F_α^2 across the image. The cruciform pattern seen in F_α^2 is due to the light profile of the input $f(\mathbf{r})$ seen in Figure 4: this fact was verified by varying the centroid position of $f(\mathbf{r})$, which was seen lead to equivalent changes in the position of the cruciform. It can be seen that $F_\alpha^2 < U_\alpha/C_\alpha$ across much of the output for this semi-realistic galaxy image.

5.2. The $\sqrt{5} \times \sqrt{5}$ Ideal Dither

In Figure 12 we plot results for five 32×32 pixel² input exposures, configured in a $\sqrt{5} \times \sqrt{5}$ ideal dither pattern. The upper-left panel, Figure 12a, illustrates this configuration of input exposures for a small sub-section of the total image area. This is an interesting dither pattern which arranges five exposures per unit cell in a slanted, but regular, grid configuration. The pattern has chirality: a reflection about the line $y = x$ produces a different, but functionally equivalent, set of dithers.

Figure 12b shows the U_α achieved for this $\sqrt{5} \times \sqrt{5}$ dither pattern: across the broad central region of the output image the requirement $U_\alpha < 10^{-8}C_\alpha$ is now successfully met. Figures 12c & d show the noise variance $\Sigma_{\alpha\alpha}$ and squared frac-

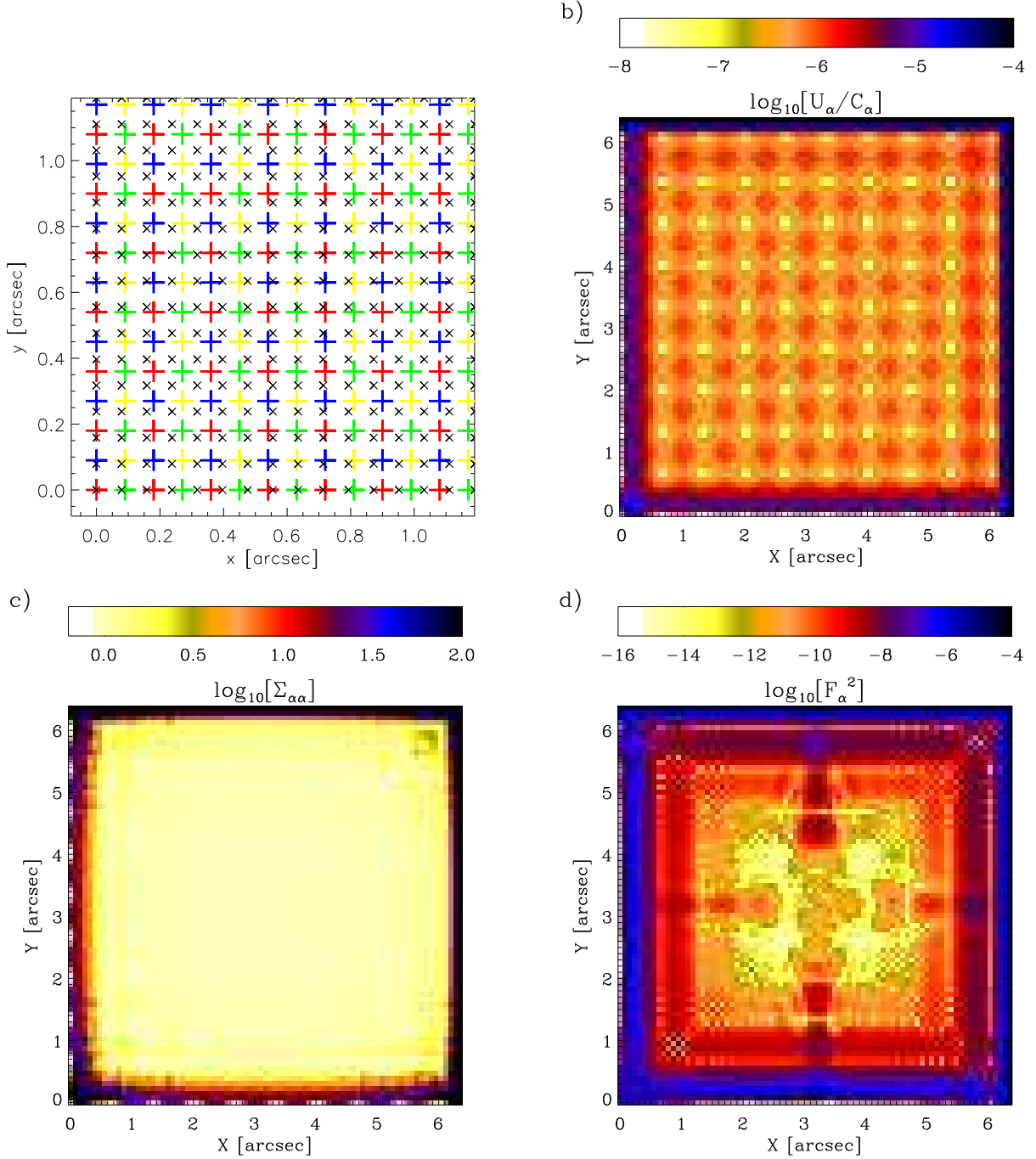


Figure 11. Input pixel configuration (a) and leakage objective U_α (b), noise variance $\Sigma_{\alpha\alpha}$ (c), and squared fractional residual F_α^2 (d) for the 2×2 ideal dither (see Section 5.1). The color range chosen for $\Sigma_{\alpha\alpha}$ does not reflect the full range of values in the output, but highlights variations in the image center.

tional residual F_α^2 . The former is seen to be approximately stationary, and comparable to the input noise variance, across the central regions of the output. The values of F_α^2 again form a vaguely-cruciform pattern that moves to align with the centroid of $f(\mathbf{r})$ when this is shifted. Overall, the $\sqrt{5} \times \sqrt{5}$ dither pattern is able to successfully reconstruct a fully-sampled output image for this *WFIRST* design concept while preserving the input convolution kernel $G_i(\mathbf{r})$ at a level $U_\alpha < 10^{-8}C_\alpha$. A question remains as to whether a pattern such as this can be used to produce fully-sampled data in the presence of image

pixel losses due to, e.g., cosmic rays, hot pixels, and other defects, or in the presence of variations in the focal plane plate scale: these issues are explored in Sections 5.4 & 5.5.

For illustrative purposes, we show in Figure 13 an example of an input exposure and the IMCOM-generated output for the $\sqrt{5} \times \sqrt{5}$ ideal dither, where actual stochastic noise η_i was added to the input pixels I_i before processing. The noise added was Gaussian with unit variance. The output H_α clearly shows the effects of increased $\Sigma_{\alpha\alpha}$ in edge regions, but demonstrates the successful image reconstruction at the

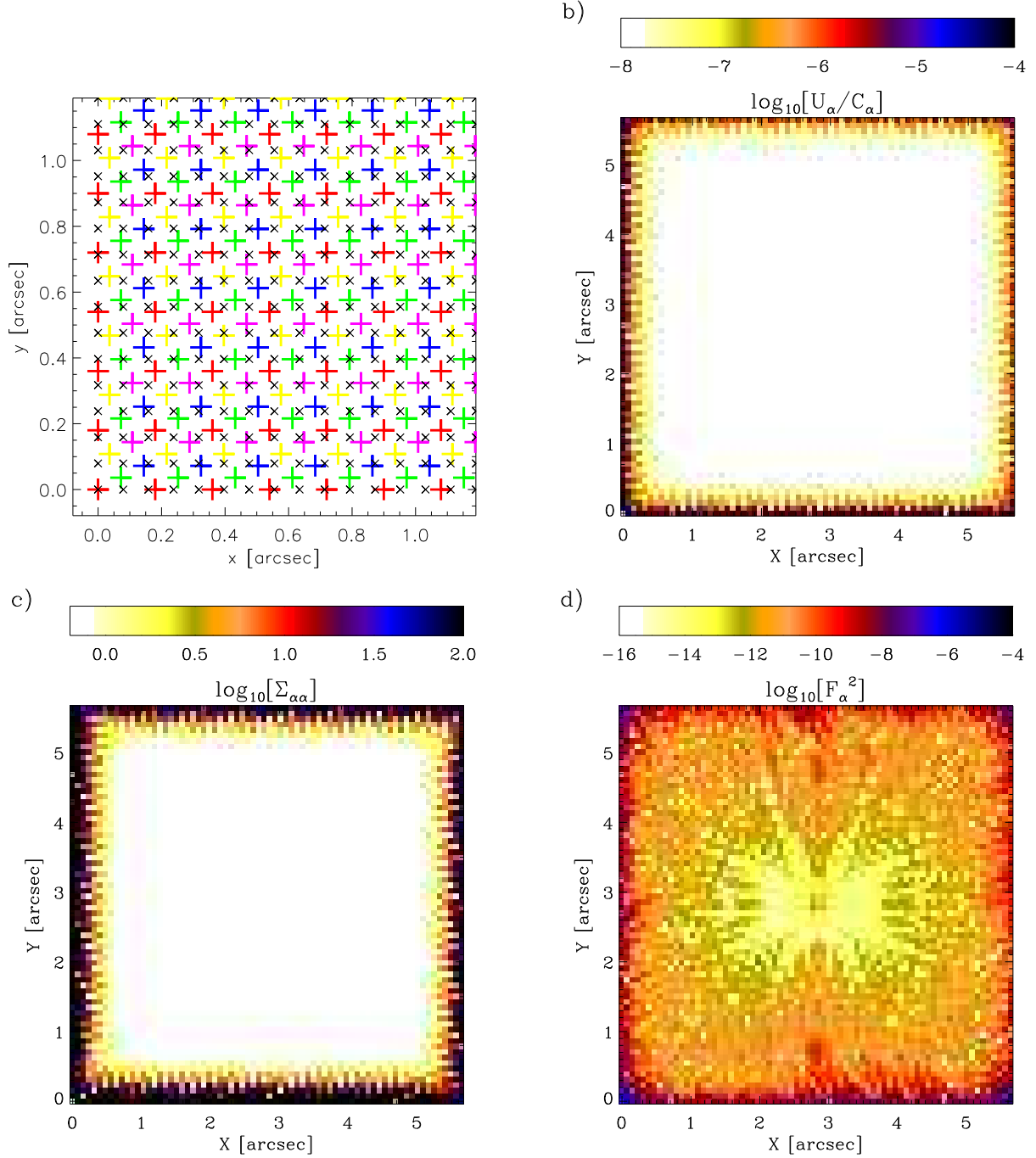


Figure 12. Input pixel configuration (a), leakage objective U_α (b), noise variance $\Sigma_{\alpha\alpha}$ (c) and squared fractional residual F_α^2 (d) for the $\sqrt{5} \times \sqrt{5}$ ideal dither (see Section 5.2). The color range chosen for $\Sigma_{\alpha\alpha}$ does not reflect the full range of values in the output, but highlights variations in the image center.

image center.

5.3. The 6 Exposure Random Dither

In Figure 14 we plot results for six 32×32 pixel² input exposures, configured in a randomly offset dither pattern; unlike in Section 4 there is no random roll in this example. Figure 14a illustrates this configuration of input exposures for a small sub-section of the total image area, along with the fully-sampling output locations. The 5 exposure random dither was also tested, but produced poor results with $U_\alpha > 10^{-8}C_\alpha$

across much of the output in a small suite of randomly offset trial configurations. One such test, not atypical of the small set of trials, produced $\Sigma_{\alpha\alpha} > 10$ for 65% of the total output and $\Sigma_{\alpha\alpha} > 100$ for 22% of the total output.

However, the leakage objective for the 6 exposure random dither in Figure 14b shows good agreement with stated requirements across a large fraction of the output sample locations. The $\Sigma_{\alpha\alpha}$ variance in Figure 14b is also now significantly smaller than that which was seen in the five exposure dithers: an impressive impact from only a single additional

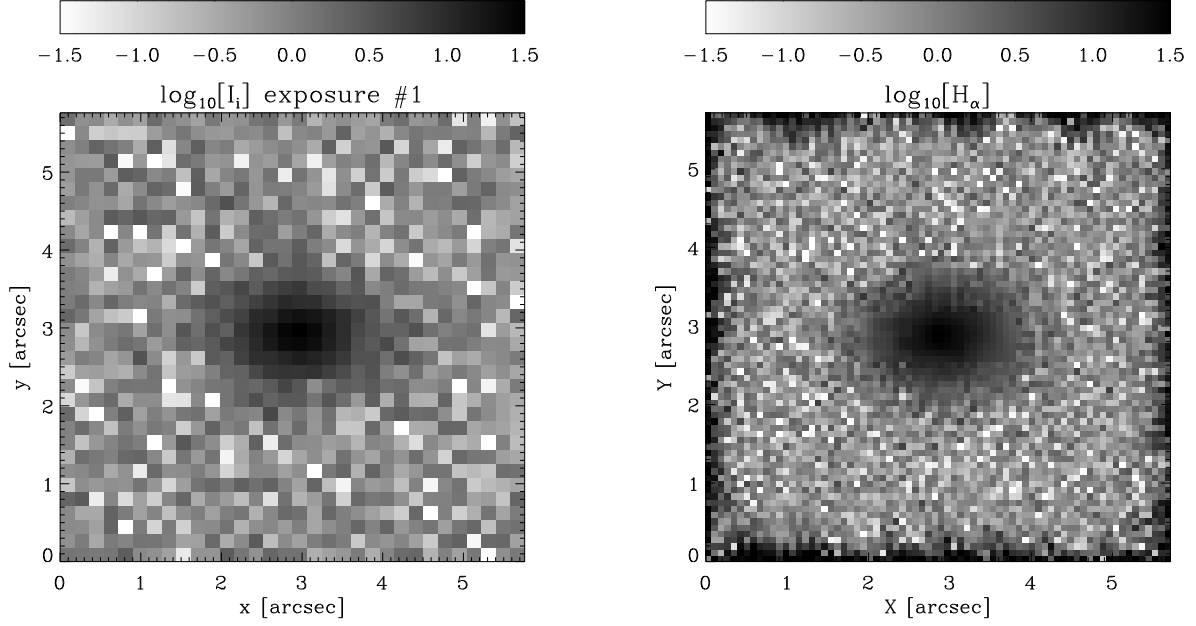


Figure 13. Left panel: One of five *noisy* input images dithered using the $\sqrt{5} \times \sqrt{5}$ ideal dither pattern of Section 5.2, constructed by convolving $f(\mathbf{r})$ with $G_i(\mathbf{r})$ and adding a stochastic noise η_i to each pixel I_i . The noise added was drawn from a Gaussian distribution with unit variance and zero mean. Right panel: The combined output image H_α . Both images show the logarithm of the absolute value of each I_i and H_α , to most clearly depict the central, edge, and background noise-dominated regions of each image.

exposure. As was done for the 5 exposure dithers, these results were repeated internally for a small set of random 6 exposure dither patterns, and the results shown are typical of the results achieved. It should be noted that the noise $\Sigma_{\alpha\alpha}$ is somewhat non-stationary, varying with \mathbf{R}_α across the output.

As has been seen, good results have been achieved for a controlled reconstruction of fully-sampled images using certain configurations of input *WFIRST*-like images. The $\sqrt{5} \times \sqrt{5}$ ideal dither performed well, as did a single realization of a random dither of six exposures (at the cost of slightly inferior noise properties). While the results shown for the random patterns were typical of a small internal test set of such patterns, the repeatability and range of these random dither results should be explored in more detail in subsequent work. Another investigation, which is within the scope of this Paper, is the robustness of linear image combination under identified image or pixel defects. This is now explored.

5.4. Cosmic Ray Impacts and Bad Pixels

We perform a very simple test of linear image combination under the presence of known image defects, such as might be caused by cosmic rays, dead pixels, and similar phenomena. Taking the input images for the $\sqrt{5} \times \sqrt{5}$ ideal dither of Section 5.2, we randomly select 5% of input pixels to be flagged as not for use in the reconstruction of the input images. This equates simply to a contraction of the $A_{\alpha ij}$ and $B_{\alpha i}$ by removal of the afflicted pixel rows and columns.

In Figure 15a we plot the map of κ_α solution values along with, in Figure 15b, the leakage objective U_α that is maintained by the algorithm for this test. Also shown are the locations of the 5% of input pixels that were thrown out, as small black points. Features in κ_α can be clearly seen where the IMCOM algorithm has attempted to satisfy requirements on U_α where input image information is scarce.

The increased noise variance in $\Sigma_{\alpha\alpha}$ for these bad pixel locations can be seen in Figure 15c. Indeed, the noise can be large in regions of missing input information, particularly where pixel removals are clustered. This information is represented in the output covariance $\Sigma_{\alpha\beta}$ given by equation (9), and is available to the user for weighting or fitting purposes if necessary. If missing pixels are expected to be clustered (such as around a single cosmic ray hit) the rapid increase in noise for such events may represent a cause for concern that should be investigated using simulations of greater realism. However, the level to which U_α (and indeed F_α^2 ; Figure 15d) may be controlled is perhaps more important for many applications, and here results are encouraging.

In Figure 16 we show equivalent results after randomly removing 5% of input pixels for the 6 exposure random dither discussed in Section 5.3. Once again, results for U_α are robust despite bad pixel losses, with $\Sigma_{\alpha\alpha}$ able to absorb uncertainty due to these small, localized instances of missing information. As for the $\sqrt{5} \times \sqrt{5}$ ideal dither, $\Sigma_{\alpha\alpha}$ may become large where missing pixels are clustered and this behavior should be investigated further.

It has been shown that the *WFIRST* design concept is able to produce output that is relatively robust to 5% pixel losses from random locations in the input images I_i , both for the $\sqrt{5} \times \sqrt{5}$ ideal dither and a 6 exposure random dither configuration. Tolerances on U_α may be maintained at the cost of increased $\Sigma_{\alpha\alpha}$ in regions of missing information: qualitatively this is the desired behavior. However, there are other practical issues that may be encountered for certain survey strategies, one of which is the possibility of significant plate scale variations between exposures when combining wide angle-dithered images. We now investigate linear image combination in such circumstances.

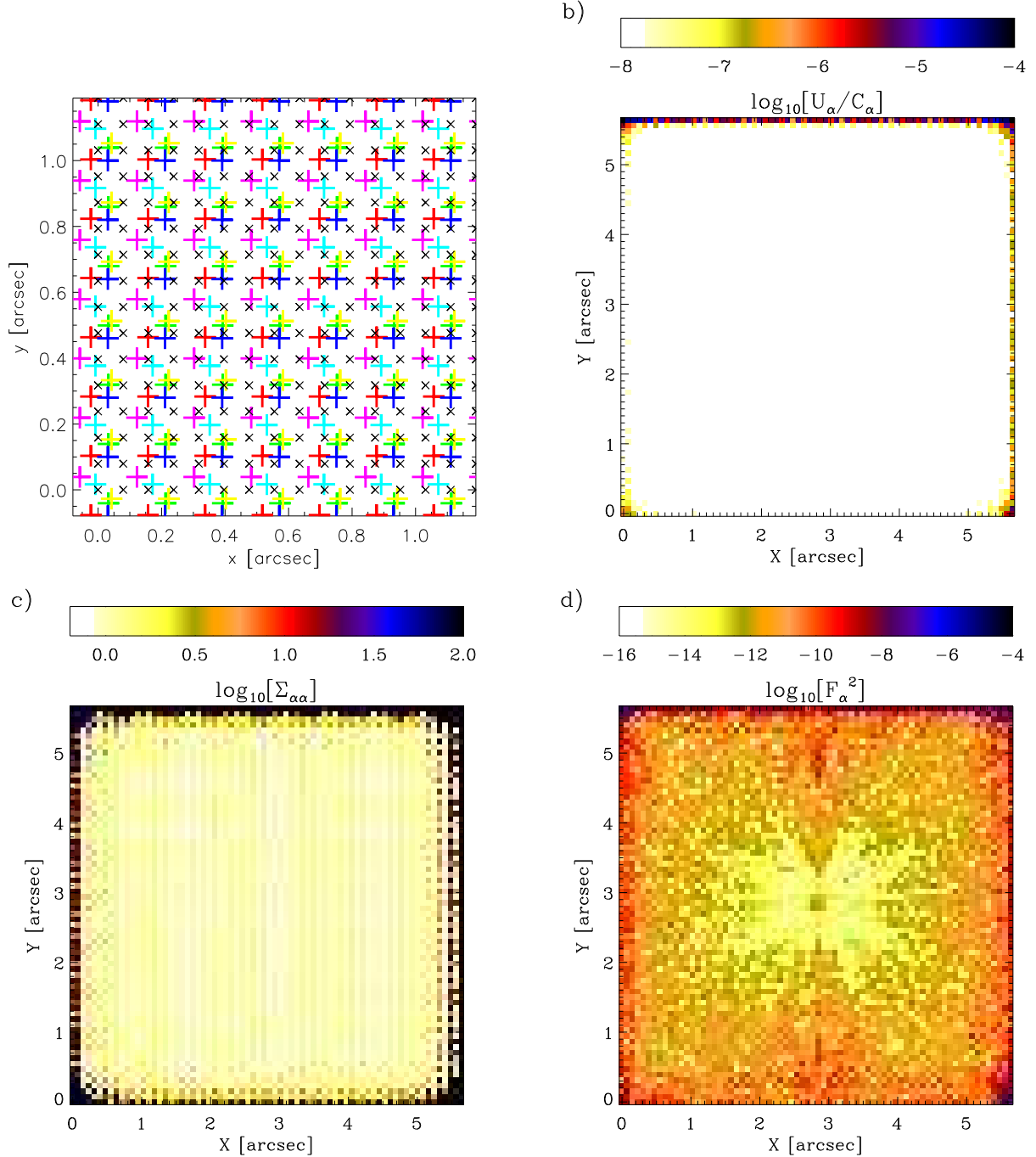


Figure 14. Input pixel configuration (a), leakage objective U_α (b), noise variance $\Sigma_{\alpha\alpha}$ (c) and squared fractional residual F_α^2 (d) for the 6 exposure random dither (see Section 5.3). The color range chosen for $\Sigma_{\alpha\alpha}$ does not reflect the full range of values in the output, but highlights variations in the image center.

5.5. Focal Plane Plate Scale Variations

Unavoidable geometric field distortions, due to optical aberration and the alignment of each detector array in the focal plane assembly, are manifested as variations in the plate scale (angular pixel scale) across a telescope field of view. These geometric distortions may be determined using precision astrometry, but it is not clear what impact they may have on a linear image reconstruction from multiple dithers. Typically the distortion is smoothly varying, so that close pairs of pixels on the detector plane show smaller relative scale vari-

ations on average than widely-separated pairs. With wide-angle slew dither strategies being considered for the *WFIRST* imaging survey, the behavior of linear image combination when facing plate scale variations in the input images is an important design consideration.

We investigate the effect of plate scale variations using the 6 exposure random dither pattern discussed in Section 5.3. For each input exposure we vary the relative fractional plate scale by a random amount drawn from a Gaussian distribution of width 1%. The variation is applied to all pixels in each

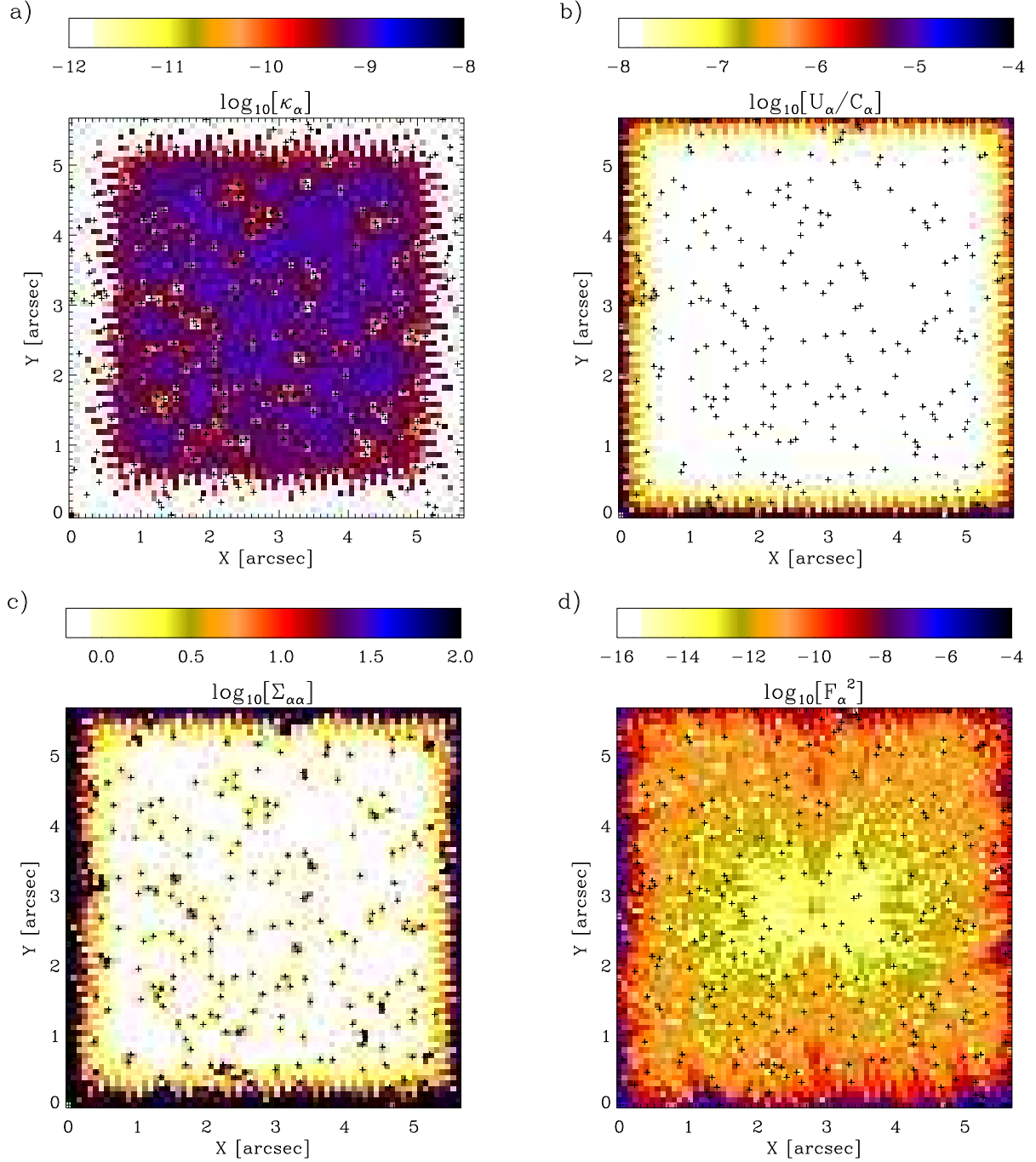


Figure 15. Map of κ_α solution (a), leakage objective U_α (b), noise variance $\Sigma_{\alpha\alpha}$ (c), and squared fractional residual F_α^2 (d) for the $\sqrt{5} \times \sqrt{5}$ ideal dither after losing a randomly-selected 5% fraction of input pixels (See Section 5.4). Input pixels that were not used are indicated using small black crosses. The color range chosen for $\Sigma_{\alpha\alpha}$ does not reflect the full range of values in the output, but highlights variations due to bad pixels in the image center.

given exposure, and is the same in both x and y directions: the effect is to uniformly expand or contract the pixel grid. The charge diffusion and pixel response component of each $G_i(\mathbf{r})$ is appropriately modified to reflect the new input pixel scale, and $\Gamma(\mathbf{r})$ is chosen as the $G_i(\mathbf{r})$ corresponding to exposure with the largest pixels.

In the upper panels of Figure 17 we plot the U_α and $\Sigma_{\alpha\alpha}$ that can be achieved for this dither configuration, having stipulated $U_\alpha^{\max} = 10^{-8}C_\alpha$ and $\Delta U_\alpha^{\max} = 10^{-10}C_\alpha$ as usual.

It is clear that the impact of $\sim 1\%$ plate scale variations is severe: $\Sigma_{\alpha\alpha}$ shows significant increases as compared to the results of Section 5.3, whereas degradation U_α is more modest but perceptible. An increase in noise variance by a factor $\gtrsim 10$ compared to the input pixels, as seen here over much of the output, is not a tolerable cost for fully-sampling a dark energy imaging survey. The success of linear image combination is clearly sensitive to small variations in the input plate scale, and a mitigating strategy must be found.

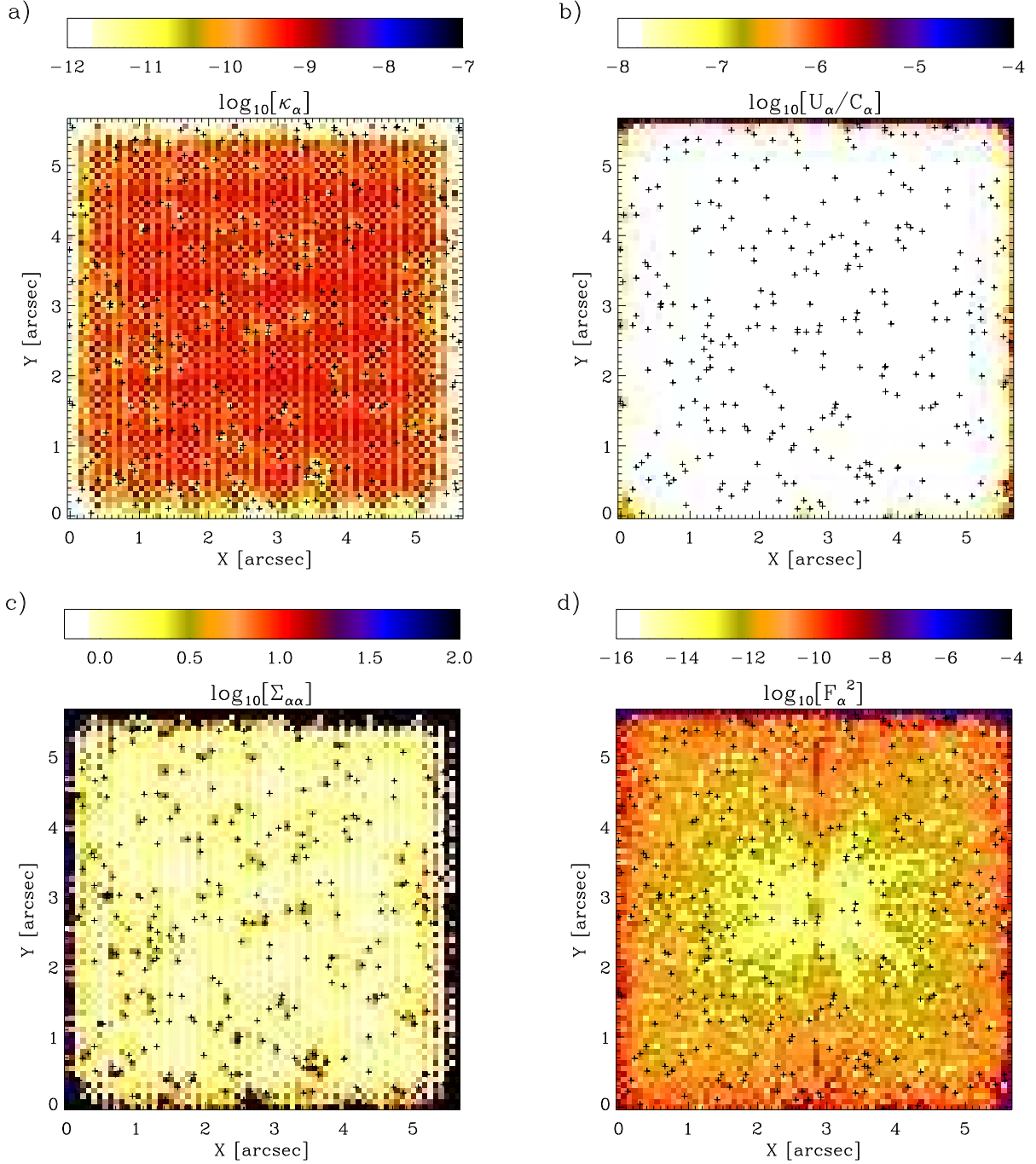


Figure 16. Map of κ_α solution (a), leakage objective U_α (b), noise variance $\Sigma_{\alpha\alpha}$ (c), and squared fractional residual F_α^2 (d) for the 6 exposure random dither after losing a randomly-selected 5% fraction of input pixels (See Section 5.4). Input pixels that were not used are indicated using small black crosses. The color range chosen for $\Sigma_{\alpha\alpha}$ does not reflect the full range of values in the output, but highlights variations due to bad pixels in the image center.

One such mitigating strategy is to increase the number of dithered exposures used in the reconstruction. This was attempted, adding additional exposures to the 6 exposure random dither, up to a total of nine. Results were modest, showing only marginal improvements in $\Sigma_{\alpha\alpha}$ as extra dithers were added.

Another option to mitigate the effects of scale variation is to exploit the user’s freedom in setting the desired output PSF $\Gamma(\mathbf{r})$. In the lower panels of Figure 17 we plot results us-

ing the same input pixel configuration as in the upper panels, but having added an additional Gaussian smoothing with $\sigma = 0.09$ arcsec (0.5 input pixels) to $\Gamma(\mathbf{r})$. Dramatic improvements to the noise properties $\Sigma_{\alpha\alpha}$ can be seen. U_α is also improved enough to bring it below $10^{-8}C_\alpha$ for a far greater proportion of the output area (see Figure 17c).

We see, therefore, that the damaging impact of plate scale variations can be suppressed to some extent by the addition of synthetic smoothing to the desired PSF $\Gamma(\mathbf{r})$. Considering the

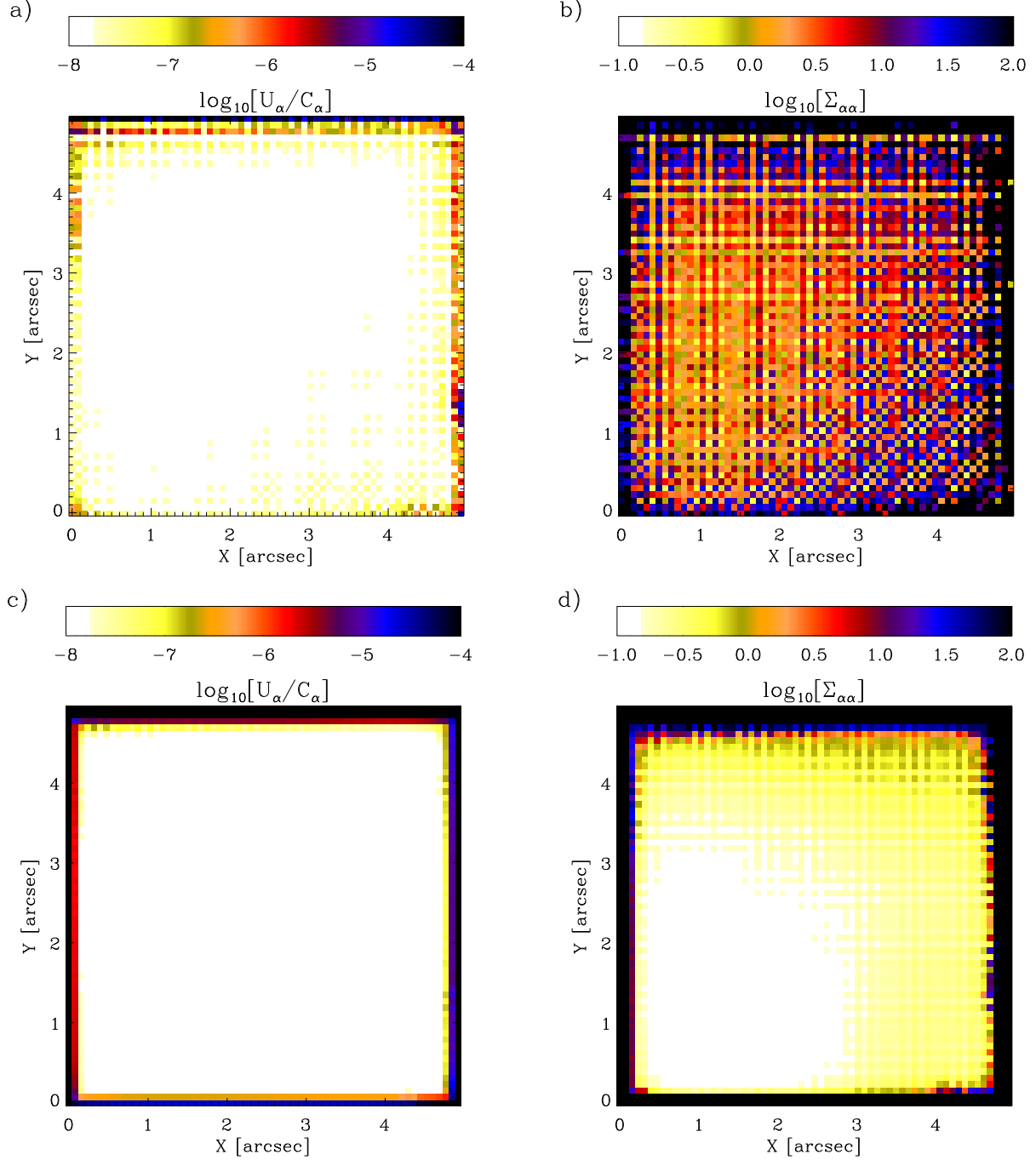


Figure 17. Upper panels: Leakage objective U_α (a), and noise variance $\Sigma_{\alpha\alpha}$ (b), for a 6 exposure random dither after adding $\sim 1\%$ variations in the pixel scale for each input exposure. Lower panels: U_α (c), and noise variance $\Sigma_{\alpha\alpha}$ (d), for the same input pixel configuration, but adding a Gaussian smoothing of $\sigma = 0.09$ arcsec (0.5 pixel) to the desired output PSF $\Gamma(\mathbf{r})$ (See Section 5.5). Plotted color scales do not necessarily reflect the full range of U_α or $\Sigma_{\alpha\alpha}$, but are chosen to emphasize the central regions of greatest interest.

effect upon the MTF $\tilde{\Gamma}(\mathbf{u})$, this can be seen simply as a suppression of the high-frequency modes that are most challenging to recover in the presence of plate scale variations. The choice of optimal smoothing filter will, in general, depend closely upon the nature of the plate scale variations encountered between input exposures, and on the dither strategy employed. The successful linear reconstruction of input images has been shown to be sensitive to this effect: further study aimed at finding optimal mitigating strategies, using more re-

alistic simulations, will be needed.

This concludes our first set of trials for the optimal linear image combination formalism. These early results suggest that the technique has merit as a design tool and can be used to explore competing dither strategies for generating over-sampled output. It has been used to identify encouraging robustness of linear reconstruction to the presence of randomly-located, missing input pixels, but has also highlighted a sensitivity to plate scale variation that should be explored further.

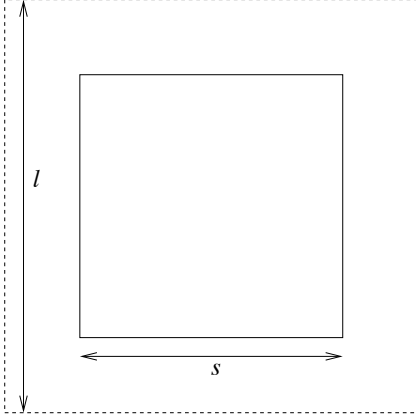


Figure 18. One approach for efficiently processing an entire dark energy survey is to split it into small patches as shown here. In this approach, input pixels I_i across multiple exposures are drawn from within a square of dimensions $l \times l$ (dashed line). From these, an output image H_α is reconstructed within a smaller, contained square of dimensions $s \times s$ (solid line). Successive patches of sky may be aligned so that the smaller output regions tessellate to cover the entire survey.

Clearly, however, an ultimate goal for the method is also its use in constructing the oversampled images from actual all-sky imaging data supplied by a mission such as *WFIRST*. Therefore, we now turn to a discussion of whether such an approach represents a computationally feasible (i.e. affordable) option for analyzing data from such a mission.

6. PRACTICAL CONSIDERATIONS & COMPUTATIONAL COSTS

The linear image combination formalism provides a natural framework for constructing oversampled images, but an equally natural question begs itself: can the method be applied to a 10 000 deg² imaging survey (such as proposed for *WFIRST*) within an acceptable timescale, and using reasonable computing resources? To answer this question, we must begin by estimating the total number of floating point operations required to apply the algorithm to a survey of this area.

As discussed in Section 3, the solution for any given patch of sky requires $O(n^3 + 2n^2m)$ floating point operations, neglecting terms that are lower order in n or m . Therefore, the optimum data processing strategy is simply to split any imaging survey into as many small patches as possible: this is clear, since n for each patch is proportional to its area. Reduced memory costs and the ease of coarse-grained parallelization are other advantages of processing many small patches of sky. In Figure 18 we illustrate one such patch schematically, noting the choice of different size input and output regions to avoid edge effects. The output regions (solid line) for such patches can in principle be laid down in a tessellating pattern to cover an entire survey area, with the larger input regions (dashed line) therefore overlapping to provide approximately uniform coverage free from the edge effects discussed in Section 5.1.

We will now consider the computational cost of such a procedure. Using the labels of Figure 18, and taking the same *WFIRST* input pixel size and output sampling rate (0.18 and 0.079333 arcsec, respectively) as Sections 4 & 5, let us consider patches of outer side $l = 25 \times 0.18$ arcsec = 4.5 arcsec and inner side $s = 38 \times 0.079333$ arcsec = 3.014 arcsec. This is a small patch, and thus an efficient means to cover a large survey, but is of sufficient size relative to the PSF to avoid significant degradation in H_α due to edge effects (see, e.g., Section 5.1). A 10 000 deg² survey requires 1.4×10^{10}

such patches for full coverage. Assuming five dithered input exposures gives $n = 3.125 \times 10^3$ and $m = 1.444 \times 10^3$ for each patch, resulting in order $n^3 + 2n^2m \simeq 5.8 \times 10^{10}$ floating point operations per patch. To process all patches independently to cover the total survey area thus requires $\sim 10^{21}$ floating point operations, a considerable computational cost.

However, because of the extreme parallelizability of the problem the use of supercomputing resources is a realistic option. The RoadRunner⁷ supercomputer at Los Alamos National Laboratory is capable of a sustained rate of 1 PFLOPS (10^{15} Floating Point Operations Per Second) in double precision arithmetic, and would therefore require approximately 12 days to complete the image combination necessary for a full dark energy survey. While it must be stressed that this is very much a “back of the envelope” estimate, and subject to usual additional factors due to practical realities and data management, the resources required are conceivably within reach even today (if costly). As of November 2010 this machine was rated as the seventh fastest on earth but, thanks in part to the advent of inexpensive Graphical Processing Units, the cost and availability of such computing power must be expected to change favorably over the next decade. Moreover, the survey imaging data also require considerable time simply to be collected by the telescope. This fact makes it unlikely that optimal linear image combination would represent a rate-determining step for a dark energy mission, even given more modest access to supercomputing resources.

Finally, this discussion has assumed that the linear combination method described in this Paper would be performed repeatedly and independently for each and every patch of sky. However, departing from this general but ‘brute force’ approach there are doubtless many gains in efficiency to be made when processing imaging data for a dark energy survey. Not least, image combination efforts might be focused in regions where significant concentrations of light have been detected on a coarser grid, avoiding empty regions of the sky. Also, there are likely to be close similarities between the PSF and input pixel configurations for nearby patches: this might allow results from costly stages in the calculation (e.g. the eigendecomposition of $A_{\alpha ij}$) to be reused multiple times. Further optimizing the implementation of the linear image combination method would make an interesting topic for future work, particularly as plans for a space-based imaging survey move closer towards reality.

7. CONCLUSIONS

It has been the intention of this Paper to demonstrate a simple, yet general, formalism for the linear combination of undersampled astronomical images to produce oversampled output. Image reconstruction based on this formalism allows the explicit control of any changes to the effective image PSF $G_i(\mathbf{r})$, and noise variance $\Sigma_{\alpha\alpha}$, due to the combination process. Such an approach is likely to be a useful addition to a growing set of image analysis tools being built up to face unprecedented challenges in astronomical inference: challenges arising primarily in the bid to understand dark energy.

Despite the numerical emphasis of the approach, we have shown how significant computational savings may be made via efficient implementation, with further savings being possible. It is planned to make the prototype Fortran 95 code IMCOM, used to generate the results of this Paper, available to the public; this will be either in its current form or as part

⁷ <http://www.lanl.gov/roadrunner/>

of a more developed, scaled-up software package for use with greater volumes of data. Even using the current prototype implementation, it has been shown that the method is a feasible means of processing all-sky data from a dark energy mission.

Within the linear framework presented, we have demonstrated some interesting results for the *WFIRST* design concept adopted in Section 4 and explored in more detail in Section 5. First, we have found that this design allows the controlled reconstruction of fully-sampled images if five input dithers (in a $\sqrt{5} \times \sqrt{5}$ ideal pattern) or six input dithers (randomly offset) are available. We also found that the $\sqrt{5} \times \sqrt{5}$ ideal dither and 6 exposure random dither configurations proved reasonably robust in U_α despite suffering 5% bad input pixel losses in random locations. Largely, these defects were able to be absorbed into an increase in the noise variance $\Sigma_{\alpha\alpha}$. Whether this conclusion is robust when the bad pixel patterns are less random remains to be seen: the contiguous spatial extent of defects such as cosmic rays is likely to be a complicating factor for the control of both U_α and $\Sigma_{\alpha\alpha}$ in the output, and merits further investigation.

Another result, discussed in Section 5.5, was the sensitivity of the linear image combination method to variations in the pixel plate scale between exposures. Plate scale variations of the order $\sim 1\%$ were seen to significantly degrade U_α and $\Sigma_{\alpha\alpha}$ for 6 exposure random dither. It was found that adding an additional Gaussian smoothing kernel of standard deviation $\sigma = 0.09$ arcsec (0.5 input pixels) to the desired output PSF $\Gamma(\mathbf{r})$ mitigated this degradation significantly. The dependence of the strength of this effect upon input image parameters such as the plate scale variation, input PSF, and dither strategy will need to be investigated more thoroughly than was possible in this Paper. Coping efficiently with plate scale variation clearly represents a non-trivial challenge when using optimal linear techniques to combine multiple input images.

In the examples tested in Section 5 it was also found that the stationarity of the noise characterized by $\Sigma_{\alpha\alpha}$ (also a desirable property if U_α may be kept low) seems to be best preserved for the ideal dither patterns, even while U_α may show strong spatial variation (e.g. Section 5.1). Conversely, setting limits on U_α for the random dither patterns causes significant spatial variation in $\Sigma_{\alpha\alpha}$. This effect is also something to be investigated further when designing dither strategies for a dark energy survey mission. Correlated, non-stationary noise introduces noise rectification biases that must be calculated and removed when performing accurate photometry and shape measurement. Precise calibration of such biases most likely proceeds best via simulations or deep (high signal-to-noise) training data, although analytic methods exist to determine the leading order terms (Kaiser 2000; Bernstein & Jarvis 2002).

For precisely such reasons, it is suggested that inference regarding galaxy shapes might best occur at the level of raw pixel images (e.g. Miller et al. 2007; Kitching et al. 2008), while image stacks or other post-processed science tools would be used to provide crucial ancillary information such as fully-sampled PSF images and precise object centroid estimates. The discussion of this question is appropriate and timely. Whatever the eventual conclusions of such a debate, there is undoubted value in a method for combining multi-

ple undersampled images to generate oversampled output that offers control over any changes to the PSF, while simultaneously suppressing noise where possible. The formalism presented in this Paper, being linear, is the simplest approach that meets these objectives.

There remain some important unanswered questions regarding the use of the technique for real data. As mentioned early in Section 2, the input PSF $G_i(\mathbf{r})$ must be known, modeled or approximated before the use of the linear image combination formalism. Any model of $G_i(\mathbf{r})$ which is sensitive to aliasing in images of stars from the input I_i risks imparting similar defects to the output image H_α . This would be a problem if attempting non-parametric reconstruction of the PSF directly from stellar images, for example. However, even in this case, a mitigating strategy may be found. If $G_i(\mathbf{r})$ does not vary rapidly across the instrumental field of view then images of different stars within a small region of a given exposure will approximate multiple images of the same PSF, but with a variety of centroid offsets (assuming variations due to star color are small, and ignoring varying flux which can be normalized). Fully-sampled images of an “average” star for each small region might thus be reconstructed by applying the linear image combination formalism in combination with an accurate estimation of the individual stellar centroids.

In practice, any model for $G_i(\mathbf{r})$ can only be an estimate of the true convolving kernel, and the impact of this uncertainty on estimates of key observables (e.g. Paulin-Henriksson et al. 2008; Rowe 2010) must be quantified in context of optimized linear image combination. A related question that must be addressed is the effect of color variation in $G_i(\mathbf{r})$ for images observed using broad-band spectral filters. Furthermore, and beyond its importance when modelling $G_i(\mathbf{r})$, the need for accurate stellar centroid estimation also arises when considering astrometric registration for each of the individual exposures that make up I_i . This registration is needed before the images may be combined, and any model of \mathbf{r}_i can only approximate the true astrometric solution. The impact of realistic levels of uncertainty in \mathbf{r}_i upon the quality of the output image H_α must be investigated. Work to explore some of these important questions is already underway in the laboratory, where they are equally relevant in the precise characterization of detector technology for a dark energy mission. Happily, the formalism presented here provides a useful framework for exploring these issues.

The authors would like to thank Tod Lauer and Gary Bernstein (particularly regarding the use of lookup tables for calculating the system matrices $A_{\alpha ij}$ and $B_{\alpha i}$) for useful discussion and suggestions, and the anonymous referee for useful comments that improved the manuscript. This work was performed in part at the Jet Propulsion Laboratory, operated by the California Institute of Technology under a contract for the National Aeronautics and Space Administration (NASA). BR has been supported by the NASA *WFIRST* Project Office. CH is supported by the Department of Energy (DE- FG03-02-ER40701), the National Science Foundation (AST-0807337), and the David & Lucile Packard Foundation.

APPENDIX

ASYMPTOTIC BEHAVIOR OF OUTPUT IMAGE PROPERTIES

The solution to the system of equations derived by minimizing W_α in equation (16), which is given by equation (20), may be written succinctly as

$$\mathbf{T} = -\frac{1}{2} (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{B}_\alpha. \quad (\text{A1})$$

Using this result with equation (9) we may write the noise variance in each output pixel as

$$\Sigma_{\alpha\alpha} = \sum_{ij} T_{\alpha i} T_{\alpha j} N_{ij} \quad (\text{A2})$$

$$= \frac{1}{4} \mathbf{B}_\alpha^T (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{N} (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{B}_\alpha. \quad (\text{A3})$$

The leakage objective U_α at each pixel may similarly be written

$$U_\alpha = \frac{1}{4} \mathbf{B}_\alpha^T (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{A}_\alpha (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{B}_\alpha - \frac{1}{2} \mathbf{B}_\alpha^T (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{B}_\alpha + C_\alpha. \quad (\text{A4})$$

Noting the simple κ dependence in the expressions above, we see immediately that

$$\frac{\partial \Sigma_{\alpha\alpha}}{\partial \kappa} = -\frac{1}{2} \mathbf{B}_\alpha^T (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{N} (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{N} (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{B}_\alpha < 0, \quad (\text{A5})$$

and

$$\frac{\partial U_\alpha}{\partial \kappa} = \frac{1}{2} \mathbf{B}_\alpha^T (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{N} \left[\mathbb{I}_n - (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{A}_\alpha \right] (\mathbf{A}_\alpha + \kappa \mathbf{N})^{-1} \mathbf{B}_\alpha > 0, \quad (\text{A6})$$

where \mathbb{I}_n is the $n \times n$ identity matrix.

These results can be interpreted in terms of the asymptotic behavior of the output image properties across the full range of κ . Taking first the limit $\kappa \rightarrow \infty$, we see that

$$\lim_{\kappa \rightarrow \infty} \{\Sigma_{\alpha\alpha}\} \rightarrow 0^+, \quad \lim_{\kappa \rightarrow \infty} \left\{ \frac{\partial \Sigma_{\alpha\alpha}}{\partial \kappa} \right\} \rightarrow 0^-. \quad (\text{A7})$$

As κ becomes large the system tends convergently to an output image of zero noise variance (or covariance). This result can in fact be trivially seen from equation (A1), as $\mathbf{T} \rightarrow 0$ in the limit $\kappa \rightarrow \infty$. The output image H_α likewise tends to zero everywhere, and the leakage and leakage objective tend as

$$\lim_{\kappa \rightarrow \infty} \{L_\alpha(\mathbf{r})\} \rightarrow -\Gamma(\mathbf{r}), \quad \lim_{\kappa \rightarrow \infty} \{U_\alpha\} \rightarrow C_\alpha, \quad \lim_{\kappa \rightarrow \infty} \left\{ \frac{\partial U_\alpha}{\partial \kappa} \right\} \rightarrow 0^+. \quad (\text{A8})$$

It follows hence that values of the leakage objective U_α are most naturally quoted in units of C_α , converging upon unity in the limit of least image fidelity.

Turning to opposing limit $\kappa \rightarrow 0$, we find that

$$\lim_{\kappa \rightarrow 0} \{U_\alpha\} \rightarrow C_\alpha - \frac{1}{4} \mathbf{B}_\alpha^T \mathbf{A}_\alpha^{-1} \mathbf{B}_\alpha, \quad \lim_{\kappa \rightarrow 0} \left\{ \frac{\partial U_\alpha}{\partial \kappa} \right\} \rightarrow 0^+. \quad (\text{A9})$$

While we can see, therefore, that the system converges stably to this minimum value there is no guarantee that $U_\alpha \rightarrow 0$ as $\kappa \rightarrow 0$, and *therefore no guarantee of an unbiased linear solution for $T_{\alpha i}$* . The noise variance exhibits the following limiting behavior

$$\lim_{\kappa \rightarrow 0} \{\Sigma_{\alpha\alpha}\} \rightarrow \frac{1}{4} \mathbf{B}_\alpha^T \mathbf{A}_\alpha^{-1} \mathbf{N} \mathbf{A}_\alpha^{-1} \mathbf{B}_\alpha, \quad \lim_{\kappa \rightarrow 0} \left\{ \frac{\partial \Sigma_{\alpha\alpha}}{\partial \kappa} \right\} \rightarrow -\frac{1}{2} \mathbf{B}_\alpha^T \mathbf{A}_\alpha^{-1} \mathbf{N} \mathbf{A}_\alpha^{-1} \mathbf{N} \mathbf{A}_\alpha^{-1} \mathbf{B}_\alpha. \quad (\text{A10})$$

Depending on $A_{\alpha ij}$ and $B_{\alpha i}$, these variances may be extremely large resulting in noisy output images. In practice a compromise between noise and fidelity will be desired.

REFERENCES

- Albrecht, A., et al. 2006, ArXiv Astrophysics e-prints: astro-ph/0609591
 —. 2009, ArXiv e-prints: astro-ph/0901.0721
 Amara, A., & Réfrégier, A. 2008, MNRAS, 391, 228
 Anderson, E., et al. 1999, LAPACK Users' Guide, 3rd edn. (Philadelphia, PA: Society for Industrial and Applied Mathematics)
 Arfken, G. B., & Weber, H. J. 2005, Mathematical methods for physicists 6th ed., ed. . W. H. J. Arfken, G. B.
 Barron, N., et al. 2007, PASP, 119, 466
 Bernstein, G. M., & Jarvis, M. 2002, AJ, 123, 583
 Blackford, L. S., et al. 2002, Transactions of the ACM on Mathematical Software, 28, 135
 Blandford, R., et al. 2010, New Worlds, New Horizons in Astronomy & Astrophysics (National Academies Press, Washington D. C.)
 Bridle, S., et al. 2010, MNRAS, 405, 2044

- Chapman, B., Jost, G., & Pas, R. v. d. 2007, *Using OpenMP: Portable Shared Memory Parallel Programming* (Scientific and Engineering Computation) (The MIT Press)
- Cheney, W., & Kincaid, D. 1998, *Numerical Mathematics and Computing*, 4th edn. (International Thomson Publishing)
- Frigo, M., & Johnson, S. G. 2005, *Proceedings of the IEEE*, 93, 216, special issue on "Program Generation, Optimization, and Platform Adaptation"
- Fruchter, A. S. 2011, ArXiv e-prints: astro-ph/1102.0292
- Fruchter, A. S., & Hook, R. N. 2002, *PASP*, 114, 144
- Gehrels, N. 2010, ArXiv e-prints: astro-ph/1008.4936
- Heymans, C., et al. 2006, *MNRAS*, 368, 1323
- Hinshaw, G., et al. 2007, *ApJS*, 170, 288
- Jarosik, N., et al. 2011, *ApJS*, 192, 14
- Kaiser, N. 2000, *ApJ*, 537, 555
- Kitching, T., et al. 2010, ArXiv e-prints: astro-ph/1009.0779
- Kitching, T. D., Miller, L., Heymans, C. E., van Waerbeke, L., & Heavens, A. F. 2008, *MNRAS*, 390, 149
- Koekemoer, A. M., Fruchter, A. S., Hook, R. N., & Hack, W. 2002, in *The 2002 HST Calibration Workshop : Hubble after the Installation of the ACS and the NICMOS Cooling System*, Proceedings of a Workshop held at the Space Telescope Science Institute, Baltimore, Maryland, October 17 and 18, 2002. Edited by Santiago Arribas, Anton Koekemoer, and Brad Whitmore. Baltimore, MD: Space Telescope Science Institute, 2002., p.337, ed. S. Arribas, A. Koekemoer, & B. Whitmore, 337–+
- Lauer, T. R. 1999a, *PASP*, 111, 227
- , 1999b, *PASP*, 111, 1434
- Ma, Z., Bernstein, G., Weinstein, A., & Sholl, M. 2008, *PASP*, 120, 1307
- Marks, R. J. 2009, *Handbook of Fourier Analysis & Its Applications*, 1st edn. (Oxford University Press)
- Massey, R., et al. 2007, *MNRAS*, 376, 13
- Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, *MNRAS*, 382, 315
- Paulin-Henriksson, S., Amara, A., Voigt, L., Refregier, A., & Bridle, S. L. 2008, *A&A*, 484, 67
- Peacock, J. A., Schneider, P., Efstathiou, G., Ellis, J. R., Leibundgut, B., Lilly, S. J., & Mellier, Y. 2006, *ESA-ESO Working Group on "Fundamental Cosmology"*, Tech. rep.
- Pence, W. D. 1999, in *Astronomical Society of the Pacific Conference Series*, Vol. 172, *Astronomical Data Analysis Software and Systems VIII*, ed. D. M. Mehringer, R. L. Plante, & D. A. Roberts, 487–+
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical recipes in FORTRAN. The art of scientific computing* (Cambridge: University Press, —c1992, 2nd ed.)
- Refregier, A., Amara, A., Kitching, T. D., Rassat, A., Scaramella, R., Weller, J., & Euclid Imaging Consortium, f. t. 2010, ArXiv e-prints: astro-ph/1001.0061
- Rowe, B. 2010, *MNRAS*, 404, 350
- Schneider, P. 2006, in *Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro*, ed. G. Meylan, P. Jetzer, P. North, P. Schneider, C. S. Kochanek, & J. Wambsganss, 269–451