TRANSIENT CORRECTION FOR ISOPHOT C-100 DETECTOR

A. Coulais[1], J. Sée[2,1], J.-F. Giovannelli[2], B. Stepnik[1], F. Balleux[1], and A. Abergel[1]

[1]Institut d'Astrophysique Spatiale, Batiment 121, Université Paris XI, 91405 Orsay cedex, France
[2]Laboratoire des Signaux et Systèmes (CNRS-SUPÉLEC-UPS), SUPÉLEC, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France.

ABSTRACT

The C-100 detector is an unstressed Ge:Ga 3×3 matrix array dedicated for IR observations between 60 to 120 μm.

The transient behavior of this detector has been studied then modeled using an empirical non linear and non symmetrical model. This non linear model has only two constant parameters. A block by block dedicated inversion method has been developed. We discuss the accuracy of this inversion method, the influence of the two parameters of the transient model, and the limitations of the method due to noise characteristics and block size.

The limitation of the photometric accuracy is estimated from the use of the correction method on the FCS observations. We also explain why the non-stationarity of the noise observed in the data indicates which data sets can be transient corrected with confidence.

Key words: ISO – PHT, C-100, infrared astronomy, infrared detectors, transients, inversion methods, cosmic rays

1. Introduction

All the low background IR photodetectors on board ISO satellite are affected by transients and memory effects (see Refs. in Coulais et al. (2000)). The amplitude of these systematic errors can be as high as 40 % for LW-CAM, 25 % for PHT C-100 and 10 % for PHT C-200.

Significant success of transient correction have been achieved for Si:Ga detectors (Schubert et al. (1995), then Coulais & Abergel (2000) and Kester (1999)) based on one of the Fouks' model (Vinokurov & Fouks, 1991).

For Ge:Ga detectors, the status of transient correction is less favorable than for Si:Ga detectors because each Ge:Ga detector seems to require a peculiar model. A non linear model was used by Caux (2001) for Ge:Ga LWS detectors. One approach based on the Si:Ga Fouks model was used for C-200 transients (Contursi et al., 2001). Recent work on C-100 (Lari et al., 2001; Rodighiero & Lari, 2001) is based on a non linear equation very closed to the Si:Ga Fouks model : both models come from the same differential Ricatti equation. From the informations given in these papers, these models have not been tested following the methodology described in Coulais & Abergel (2000) and in Coulais et al. (2001) and based on steps of incoming flux. As a consequence it is very difficult to reach any conclusion (1) on the adequation of the model with the data, (2) on the validity range of the model and (3) on the final accuracy. For LWS and SWS, it is quite impossible to use this methodology because the spectroscopic observations are made by continuous change of wavelength then we do not have useful steps of flux. This study can be done for C-100 and C-200 with PHT P22 observation mode.

Since it is often very difficult to re-use for another detector these transient models and the associated inversion methods, full details must be given on the selected data, on the chosen model, on the known limitations of the model, and on the inversion methods. It is necessary to provide full software which include not only the inversion method but also the direct model (simulation tools).

We have briefly described in Coulais et al. (2000) which exponential but non linear model we have chosen for C-100. In this paper we detail this work. In Sect. 2, we present the selected data and the developed model. Model validity is discussed. The block by block inversion method is described in Sect. 3. In Sect. 4, we discuss the main problems and limitations we have encountered concerning the C-100 data processing. We also discuss the accuracy achieved before and after this transient correction.

2. Data and model selection

2.1. Detector and observation mode

PHT C-100 is a 3×3 matrix array in one Ge:Ga bulk. C-100 covers mid-infrared domain between 60 and 120 μm (Lemke et al., 1996). One pixel (#8) has always a S/N ratio two times lower than the other ones.

We use the Gabriel & Acosta-Pulido (2000) terminology : each non-destructive measurement in the ramp is called *readout*, each value computed from the slope of one ramp between two destructive measurements is called *signal*.

We consider only P22 observations; nevertheless this work should be used for any observing mode for the same detector since transient response remains the same. C-100 observations in P22 mode are based on pointed observations. One pointed direction corresponds to one block (incoming flux is the same during this block). During one pointing, tens to hundred signals are computed from typically 8 to 256 readouts. One signal is the result of the ramp processing (Gabriel & Acosta-Pulido, 2000). Sampling rate and ramp length were selected following expected brightness of the observed regions. We call inter-block the set of signals between two successive pointed observations. The incoming flux changes during all the inter-blocks.

2.2. Data selection

From previous works on such Ge:Ga detectors (Swinyard et al., 1996) and from theory (Fouks, 1992; Haegel et al., 1999), it is expected that the transient response is a non linear one. Characterizing such a non linear model is not trivial even when an adequate model is available. One example of this kind of work is detailed in Coulais & Abergel (2000). In Coulais et al. (2001) we have explained that it is much more convenient to work on test signals based on steps of flux. At the beginning, it is better to work on flux steps with a stabilized initial level in order to avoid memory effects[1]. If needed, any analytical model can be modified to work on continuously variable signal.

To select the data needed to characterize the transient several criteria are important in order to simplify the study : (1) the inter-block time should be as short as possible, (2) both upward and downward steps of flux are needed since the response is generally not symmetrical, (3) the whole range of initial and final levels should be covered in order to constrain any model, (4) the memory effect (which should not be confused with transient effect) should be as low as possible.

Under non-uniform illumination of the pixel surface or of adjacent pixels, contrary to LW-CAM matrix array (Coulais & Fouks, 2001), it is not evident from data that C-100 is affected by electrical crosstalk between adjacent pixels. When the illumination gradient is high between the adjacent pixels, we do not see a change of transient shape (no overshoot as observed with LW-CAM) and model parameters β and λ (see Sect. 2.3) do not seems to change. As a consequence, we consider the pixels are independent.

We have used asteroid observations[2] for model characterization because they contain upward and downward steps with different initial and final levels and the blocks are long (hundreds of signals). One drawback of these data is that the inter-block is long (tens of signals). Limitations are that (1) the initial level is generally the same for the upward steps and (2) the final level is generally the same for the downward steps (these levels correspond to the observed background level of the sky). Moreover the total number of upward and downward steps is limited to $\sim$ ten steps for each pixel (but few steps with high level).

2.3. Model selection

A first look at the selected data clearly shows that this kind of transient cannot be described by the Fouks model (Vinokurov & Fouks, 1991; Fouks & Schubert, 1995) used for ISO Si:Ga detectors[3], which is not surprising, since the assumptions used to simplify the physical equations are made for Si:Ga detectors. C-100 is a Ge:Ga one[4].

From the data, the C-100 transients characteristics can be described as follow:

- upward incoming flux step results, after an instantaneous jump in an exponential increase of the signal up to a stabilized level,
- contrary to transients for Si:Ga detectors, no inflection point have been observed in the upward steps,
- the higher the amplitude of the upward step the faster the stabilization[5],
- downward incoming flux step results in an instantaneous decrease of the signal down to a stabilized level,
- the memory effect is reseted when a downward step occur.

It should be noticed that, because of the length of the inter-block time, no "instantaneous" behavior is really observed in the data. Due to the limited number of usable steps of flux, it was also very difficult to work properly on the time constant characterization and on the memory effect.

[1] Here appears the subtle difference between transient response and memory effect. When a step of flux occurs at time t_0 after a *stabilization* of the detector, the transient response is not affected by memory effects. But when the stabilization is not reached for the block $n-1$, the transient response departs from the stabilized one. How to define this *stabilization*? In the Fouks' model used for Si:Ga detectors the memory effect is synthesized in only one value ($J_{n-1}(t)$) and we need only three values ($J_{n-1}(t_0^-)$, J_{n-1}^{∞} and J_n^{∞}) to compute any $J_n(t > t_0)$. Then we can said that the stabilization is achieved when $J_{n-1}(t)$ is very close to J_{n-1}^{∞} ("in the noise"). This criterion could be insufficient for Ge:Ga detectors : it is predicted from physical theories (Fouks, 1996, 1997) that the memory effect cannot be summarized in only one value for these detectors. The physical origin of this difference lies in the fact that the applied electrical field is much lower for Ge:Ga than Si:Ga. Then the knowledge of the whole prehistory may be needed to predict the transient response after a new step of flux for some Ge:Ga detectors. This effect appears clearly on ground based tests of ASTRO-F Ge:Ga detectors (Kaneda et al., 2001). No evident example were founded for Ge:Ga detectors on-board ISO.

[2] Partial list of TDT : 08600902, 08601205, ..., 26503102, 26503405, ..., 27500905, 27501208, ...

[3] Technics used for characterization of Si:Ga Fouks model were described in Coulais & Abergel (2000).

[4] More precisely the Si:Ga Fouks model is derived assuming a ratio $E_0/E_j \gg 1$, where E_0 is the applied voltage and E_j is a parameter which indicate the contacts quality of the detector. This is true for Si:Ga detectors. For Ge:Ga, $E_0/E_j \ll 1$.

[5] From the limited number of upward steps, it was not possible to choice between the next two possibilities : the time constant is related to the amplitude of the current step or to the difference of the amplitudes of the current step and the previous one. We have decided to use the simplest one.

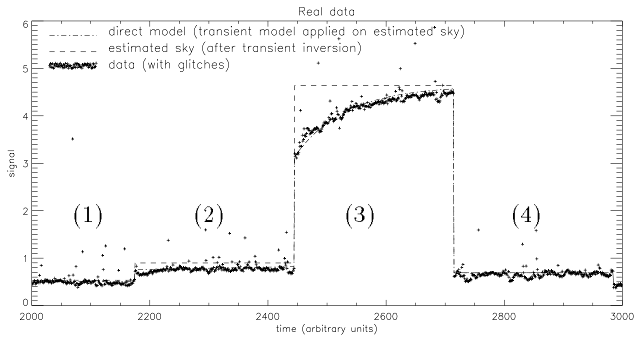


Figure 1. Example of transient response and transient correction on experimental data for an high upward step. First block (t between ∼2000 and 2175) is assumed to be stable. Second block (t between ∼2175 and 2450) shows a long term transient on this small upward step. Third block (t between ∼2450 and 2710) shows a fast transient, this transient is close to stabilization at the end of the block. Since the time constant is proportional to λ/J^∞, higher the flux step, faster the stabilization. Fourth block (after t ∼2710) shows an quasi instantaneous downward step. On these data the inter-blocks have been removed. Flux estimations and output of the transient model on the flux estimations are also over-plotted.

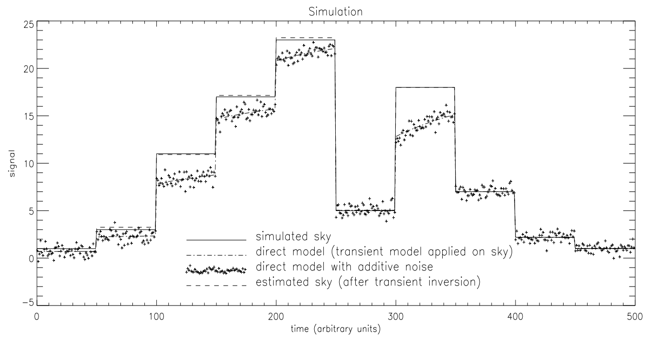


Figure 2. Example of transient correction on simulated data. Incoming flux is composed by a series of constant blocks. Blocks size is 48 (possible blocks sizes are between 8 and 256, by factor 2). Out coming flux is simulated using the direct model plus Gaussian noise, with strong amplitude compared to real data. Except for blocks 50-100 and 200-250, transient correction is perfectly close to the initial incoming flux.

Several empirical models have been tested, initially based on exponential term. It becomes evident that the exponential term should include the current illumination level (Sée, 1999). After testing different models (linear and non linear, empirical or physical from Fouks' team (Fouks, 1992)), we have concluded that a very simple empirical non linear model with a non symmetrical behavior can reproduce the observations :

$$J_n(t) = J_n^\infty - (1-\beta)\,\varphi(J_n^\infty - J_{n-1}^\infty)\,e^{-J_n^\infty (t-t_n)/\lambda}, \qquad (1)$$

where t is the time, n the block number, $J_n(t)$ the measured signal at instant t (after ramp processing), J_n^∞ the incoming flux observed during the block n. $\varphi(u) = 0$ if $u \leq 0$ and $\varphi(u) = u$ if $u > 0$. This model has only two parameters : β the instantaneous jump and λ a constant (the time constant is λ/J_n^∞).

It should be noticed that, for the upward steps, this model was described by Fouks (1992). We have selected a model which is different than the model used by Acosta-Pulido et al. (2000).

The way used to estimate the two parameters (β, λ) is described in the next Sect. 2.4. The dedicated inversion method is described in Sect. 3.

2.4. Parameters estimation

The transient model (Eq. 1) is characterized by two parameters (β, λ). These parameters are supposed to be constant for one given pixel. They must be estimated one time for each pixel before using the inversion method. We used upward steps of flux to fit simultaneously β and λ. Major technical problem in fitting procedure comes from the inter-blocks. It had been chosen to remove these inter-blocks before fitting.

Parameters (β, λ) have been fitted with Eq. 1 for a limited number of upward steps covering the range from $J_1^\infty/J_0^\infty \sim 2$ to $J_1^\infty/J_0^\infty \sim 20$ ($J_0^\infty \leq 1\ V$) : $\beta = 0.7 \pm 0.1$ and $\lambda = 100 \pm 40$. We found a dispersion of 15% for β and 40% for λ for the upward steps in our database. These rather high dispersions come from two problems : (1) the number of upward steps we used is too small to strongly constrain β and λ values; and (2) glitches clearly appear in the data and change the transient response during one block. Then we think that inter-block removing is not the critical point at this time.

We have assumed that β and λ are constant for one given pixel. At the present time, we don't have analyzed enough data in order to study the possible variations of these two parameters during the whole ISO mission. Furthermore it is known that C-100 responsivity increase by few percents between curing procedures (Lemke et al., 1996; Wilke et al., 2001).

3. Block by block inversion method

The method was described from a more general point of view in Coulais et al. (2000). In this paper, it was also explained how to adapt this block by block inversion method for others models.

This inversion method is a block by block method in the sense that only one incoming flux value has to be estimated for each block. The main hypotheses are : (1) the pixels are independent, (2) for each pixel the detector parameters (β, λ) are known and are positive constant, (3) for each pixel the incoming flux is constant during one block[6].

[6] The inter-block can be processed as a series of blocks with one signal each. This way has the drawback to strongly increase the total computing time.

We have to invert the model $\mathcal{M}$ (see Eq. 1) and we have N data blocks, with indices n ($n = 1, \ldots, N$). Measured signals in one block are M samples of $J_n(t)$ gathered in the $\boldsymbol{J} \in {}^M$ vector. The N unknown fluxes (one per block) are J_n^∞, $n = 1, \ldots, N$, gathered in the vector $\boldsymbol{J}^\infty \in {}^N$.

The inversion, i.e. the estimation of $\boldsymbol{J}^\infty$ from $\boldsymbol{J}$, is achieved by least square method. Let Q_{LS} be the least square distance between experimental data $\boldsymbol{J}$ and outputs $\mathcal{M}(\boldsymbol{J}^\infty)$ of the model to be inverted :

$$Q_{\mathrm{LS}}(\boldsymbol{J}^\infty) = \| \, \boldsymbol{J} - \mathcal{M}(\boldsymbol{J}^\infty) \, \|^2. \tag{2}$$

We have to compute the value of the unknown object $\boldsymbol{J}^\infty$ which minimizes Q_{LS} :

$$\widehat{\boldsymbol{J}}_{\mathrm{LS}}^\infty = \arg\min_{+} Q_{\mathrm{LS}}(\boldsymbol{J}^\infty) \, . \tag{3}$$

By this way, we simultaneously use the two available kinds of information : the unknown flux is positive and is constant by blocks.

Criterion optimization is the critical part of this work, mainly because of the non linear and non symmetrical nature of model $\mathcal{M}$. Nevertheless Q_{LS} has a peculiar structure because, for each block, the model depends only by two unknown values J_n^∞ and J_{n-1}^∞. Using this property, we can write the least square criterion Q_{LS} as follow :

$$Q_{\mathrm{LS}}(\boldsymbol{J}^\infty) = \sum_{n=2}^{N-1} Q_n(J_n^\infty \, , \, J_{n-1}^\infty) \, , \tag{4}$$

which shows a coupled structure. This kind of structure is typical from Markov chains (Rabiner & Juang, 1986). Global minimization is then possible on a discrete grid with K incoming flux values using Viterbi algorithm (Forney, 1973).

This is a recursive algorithm on the n blocks. In a first step, this algorithm computes the least square criterion for all the possible values of (J_1^∞,J_2^∞) couple on the grid. It extracts for each J_2^∞ the minimum with respect to J_1^∞ and retains the K minima and K minimizers. Next step introduces J_3^∞. From first step results, the algorithm minimizes in respect with J_2^∞ for all possible J_3^∞ values. Then, step by step, the optimization is done till J_N^∞. The last operation consists in step by step back tracking in order to find the minimizers J_{N-1}^∞, J_{N-2}^∞ till J_1^∞.

The grid can be adjusted in order to improve the optimization. The local optimum could be improved using a gradient method.

Example of transient correction on simulated data is shown on Fig. 2. Despite a high amplitude Gaussian noise compared to experimental noise, transient correction is better than 1% level, even with successive upward steps. This accuracy is achieved even for small block size (32). As a direct consequence, the accuracy of the inversion method is not a limitation. The problems should come from unadapted model, non Gaussian noise, ...

Example of transient correction on experimental data is shown in Fig. 1. The figure also shows the modeled transient and the estimated values (constant by block).

4. Known limitations and discussions

We have summarized the main problems in this section, in order to simplify the description of the direct problem and the inversion method. The data have been processed with PIA V 7.2.2. Deflating and mapping problems and algorithms are not addressed here.

4.1. Glitches and noise study

Deglitching methods for PHT detectors were described in Gabriel & Acosta-Pulido (2000). In this paper, Fig. 8 and Table I show that C-100 was significantly affected by glitches. A very interesting example of C-100 transient was also presented in Fig. 7 in Acosta-Pulido et al. (2000). This figure justifies also that we can work with a rather high dispersion for the (β, λ) parameters : how to define the *true* transient response in this signal clearly affected by glitches and small oscillations ?

As recalled in Gabriel & Acosta-Pulido (2000), “for a constant illumination the measured signals should approximately form a Gaussian distribution, whereas glitches show up as tails in such distributions”. As expected, we clearly see in the noise distributions[7] that the tail significantly increases when the readout frequency decreases (the faster the readout frequency, the easier the glitch correction at ERD level). As a direct consequence, the data with low readout frequency cannot be processed with high confidence. Furthermore, data with block sizes of 8 or 16 cannot be transient corrected with high confidence.

We illustrate on Fig. 3 the degradation of noise gaussianity with the increase of integration time in ramps. A possible way to account for heavy-tailed statistics relies on robust statistics, e.g. replace the L_2 norm used in Eq. (2) by a L_1 one.

4.2. About FCS and photometry

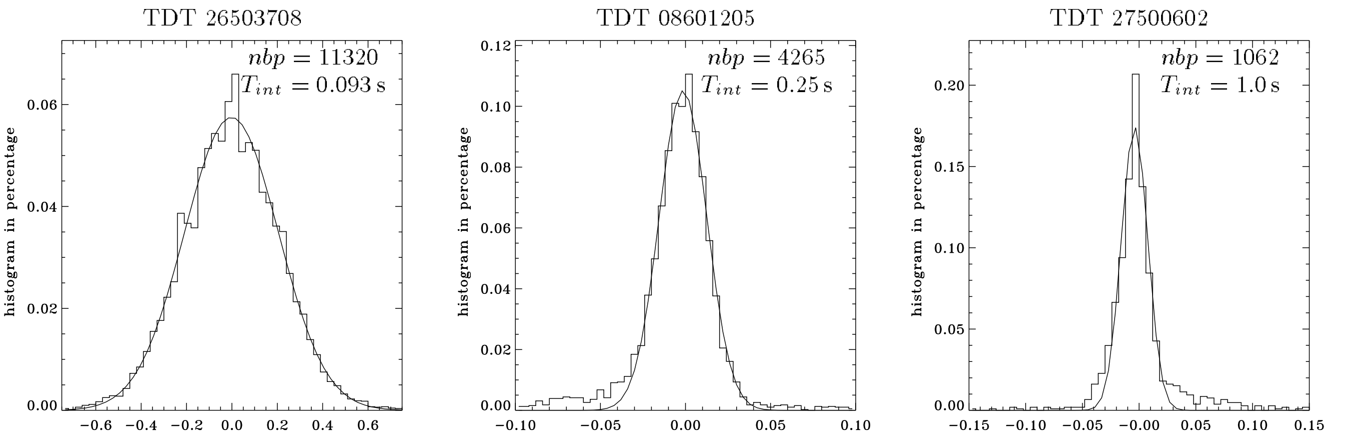
Before and after each P22 scientific observation, FCS[8] have been observed (FCS-1 before, FCS-2 after). These FCS are used to derive the photometry. FCS are also affected by transient. Accurate transient correction of the FCS is then critical to obtain an accurate photometry.

After studying these FCSs we think that they have several drawbacks. Generally, we don't know the prehistory of FCS-1. As a consequence, transient correction method cannot be used directly on them. It has been decided to use only the FCS-2. Moreover FCS observations are generally too short and too noisy to be corrected with an high accuracy (the accuracy needed on FCS is greater than for observations since FCS are related to absolute photometry).

---

[7] During this noise study, a possible problem of signal quantification was exhibited. We plane to reprocess the data with the last PIA version to confirm this possible problem.

[8] internal Fine Calibration Sources.



)*Figure 3. Noise distribution for three data sets with different integration time T_{int} and different amount of samples nbp (TDT 26503708 : $T_{int} = 0.09375$ s., nbp = 11320; TDT 08601205 : $T_{int} = 0.25$ s., nbp = 4265; TDT 27500602 : $T_{int} = 1.$ s., nbp = 1062). Those distributions are computed removing a moving mean with width 7 in the signal data. The non-Gaussian noise is clearly visible on the distributions for TDT 08601205 and 27500602. Transient correction must not be apply on these data.*

Depending on the ramp parameters (number of points and readout duration) and the number of signals in FCS, we estimate to $\sim$ 5 to 15% the error on standard FCS-2 after transient correction (resp. $\sim$15 to 25% without transient correction). One part is coming from the noise level, another part comes from hysteresis (as defined in Gabriel & Acosta-Pulido (2000)) induced by glitches and last part comes from the finite length of the FCS.

5. Conclusion

Transient response of PHT C-100 have been studied then modeled. A non linear model with only two constant parameters was chosen. A block by block inversion method has been developed. This method can be transposed for other models. During this study it becomes clear that available data did not allow a full study of transient response and memory effect. When the integration time is short enough, this transient correction method is useful to reduce remanence (bias) in raster maps. Local differences up to 15% in flux have been found in maps before and after transient correction (e.g. TDT 31800304).

We think that now the limitations in photometry accuracy for PHT C-100 did not come from the errors introduced by the transient response but come from the limited accuracy obtain from FCS and the bias introduced in signal by the rather high sensitivity of PHT C-100 to glitches. We think that these two problems actually reduce the usefulness of any transient correction methods for PHT C-100.

ACKNOWLEDGEMENTS

AC thanks CNES, CEA SaP and IAS for financial support. AC thanks him students J. Sée, F. Balleux and S. Bouaissi who have worked hard on this topic. AC thanks G. Lagache and B. Stepnik who have preprocessed the data. AC thanks G. Lagache who provides a deflating program.

References

Acosta-Pulido, J. A., Gabriel, C., & Castañeda, H. O. 2000, Experimental Astronomy, 10, 333
Caux, E. 2001, in this volume, E47–+
Contursi, A., Boselli, A., Gavazzi, G., Bertagna, E., Tuffs, R., & Lequeux, J. 2001, A&A, 365, 11
Coulais, A. & Abergel, A. 2000, A&A S.S., 141, 533
Coulais, A., Abergel, A., Pérault, M., & Fouks, B. 2001, in this volume, E45–+
Coulais, A. & Fouks, B. 2001, in this volume, E73–+
Coulais, A. et al. 2000, Proc. of SPIE, 4131, 205
Forney, G. D. 1973, Proc. of the IEEE, 61, 268
Fouks, B. I. 1992, in ESA SP-356, 167
Fouks, B. I. 1996, Proc. of SPIE, 2817, 160
Fouks, B. I. 1997, Proc. of SPIE, 3122, 441
Fouks, B. I. & Schubert, J. 1995, Proc. of SPIE, 2475, 487
Gabriel, C. & Acosta-Pulido, J. A. 2000, Experimental Astronomy, 10, 319
Haegel, N. M., Simoes, J. C., White, A. M., & Beeman, J. W. 1999, Applied Optics, 38, 1910
Kaneda, H. et al. 2001, in this volume, E78–+
Kester, D. 1999, The impact of memory effects correction on SWS data, Tech. rep., SRON, Netherlands
Lari, C. et al. 2001, accepted in MNRAS
Lemke, D. et al. 1996, A&A S.S., 315, L64
Rabiner, L. R. & Juang, B. H. 1986, IEEE ASSP, 4
Rodighiero, G. & Lari, C. 2001, in this volume, E56–+
Schubert, J., Fouks, B. I., Lemke, D., & Wolf, J. 1995, Proc. of SPIE, 2553, 461
Sée, J. 1999, Correction des effets mémoire de l'instrument ISOPHOT C-100, Tech. rep., Stage de magistère
Swinyard, B. M. et al. 1996, A&A S.S., 315, L43
Vinokurov, L. A. & Fouks, B. I. 1991, Sov. Phys. Semicond., 25, 1207
Wilke, K., Grözinger, U., Klaas, U., & Lemke, D. 2001, in this volume, E49–+