



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Σχολή Θετικών Επιστημών – Τμήμα Φυσικής
Τομέας Αστροφυσικής, Αστρονομίας και Μηχανικής

Πτυχιακή εργασία

**Χρήση αλγορίθμων FFT και Fast Folding στην
ανίχνευση ασθενών περιοδικών αστρονομικών σημάτων**

Search for weak periodicities in astronomical data using the FFT and FFA algorithms

Συγγραφέας: Κωνσταντίνος Κ. Κοβλακάς
Επιβλέπων: Καθ. Ιωάννης Χ. Σειραδάκης

Τριμελής επιτροπή

Καθ. Ιωάννης Σειραδάκης
Καθ. Εμμανουήλ Πλειώνης
Αναπλ. Καθ. Νικόλαος Στεργιούλας

Θεσσαλονίκη, Μάιος 2013

Στην οικογένειά μου,
για την συνεχή υποστήριξη και ενθάρρυνση

Περίληψη

Στις μέρες μας, τα επίγεια και διαστημικά τηλεσκόπια συμμετέχουν σε έρευνες χιλιάδων ή ακόμα και εκατομμυρίων ουρανίων σωμάτων. Συχνά, ο στόχος τέτοιων δραστηριοτήτων είναι ο χαρακτηρισμός / ταξινόμηση πολλαπλών στόχων με την ανίχνευση περιοδικών φαινομένων ενδογενούς (π.χ. Κηφείδες μεταβλητοί) ή εξωγενούς χαρακτήρα (π.χ. διαβάσεις εξωπλανητών), στις καμπύλες φωτός τους. Η επεξεργασία πολυάριθμων μετρήσεων για καθέναν από τους στόχους είναι μία χρονοβόρα διαδικασία και η χρήση ‘έξυπνων’ αλγορίθμων κρίνεται απαραίτητη.

Ο Γρήγορος Μετασχηματισμός Fourier (FFT)) είναι μία γνωστή μέθοδος για την ανάλυση χρονοσειρών και χρησιμοποιείται κατά κόρον σε έρευνες. Ωστόσο, σε περιπτώσεις μη ημιτονοειδών κυματομορφών (π.χ. pulsar), ο FFT μπορεί να φανεί ασθενής. Μία εναλλακτική μέθοδος, ο Fast Folding Algorithm (FFA), εισήχθη από τον David Staelin (Staelin, 1969) με σκοπό να λύσει αυτό το πρόβλημα.

Δημιουργήσαμε πραγματοποιήσεις και των δύο αλγορίθμων, τις οποίες ελέγξαμε για ακριβή έξοδο και ταχύτητα. Έπειτα, τις χρησιμοποιήσαμε για να εκτιμήσουμε την περίοδο σε χρονοσειρές διαφόρων παραμέτρων (περίοδος, θόρυβος κτλ.) που προσομοιώνουν τις καμπύλες φωτός διαφόρων τύπων αστρονομικών αντικειμένων και συγκρίναμε τα αποτελέσματα. Επίσης, αναλύθηκαν και πραγματικά αστρονομικά δεδομένα από την αποστολή Kepler της NASA.

Επιβεβαιώσαμε την καλή λειτουργία των αλγορίθμων, καθώς και την προτεινόμενη υπολογιστική πολυπλοκότητα των FFT (από τη βιβλιογραφία) και FFA (από εμάς). Βρήκαμε πως και οι δύο αλγόριθμοι ανιχνεύουν τη σωστή περίοδο στις περισσότερες των περιπτώσεων. Δυστυχώς, η επεξεργασία της εξόδου των αλγορίθμων αποδείχθηκε μάλλον απλοϊκή και απέτυχε να αναδείξει την υπεροχή

του FFA έναντι του FFT για περιοδικούς παλμούς μικρού κύκλου εργασίας.

Τέλος, προτείνουμε διορθώσεις του κώδικα και μία μέθοδο αναζήτησης περιοδικοτήτων που συνδυάζει τους δύο αλγόριθμους.

Ευχαριστίες

Θα ήθελα να εκφράσω την ευγνωμοσύνη μου στον επιβλέποντα καθηγητή και εμπνευσμένου διδασκάλου, Ιωάννη Σειραδάκη¹ για τις χρήσιμες κατευθύνσεις και το ειλικρινές ενδιαφέρον που επέδειξε για την εργασία αυτή.

Θέλω να ευχαριστήσω τους προσωπικούς μου φίλους, Γιάννη Μαδεμλή² για τις συμβουλές του πάνω σε θέματα Πληροφορικής και επεξεργασίας σήματος, και Έλλη Τζώρτζογλου³ για την πρόθυμη βοήθειά της στη διόρθωση του κειμένου.

Τέλος, οφείλω ένα μεγάλο ευχαριστώ στη Χριστίνα Νούτσου για την υπομονή και την κατανόησή της τους τελευταίους μήνες.

Η εργασία αυτή έκανε χρήση των

- NASA Exoplanet Archive, που λειτουργείται από το California Institute of Technology, βάσει σύμβασης με το Πρόγραμμα Εξερεύνησης Εξωπλανητών της Εθνικής Υπηρεσίας Αεροναυτικής και Διαστήματος των Ηνωμένων Πολιτειών της Αμερικής
- βάση δεδομένων SIMBAD, που λειτουργείται από το CDS, Στρασβούργο, Γαλλία
- Astrophysics Data System της NASA (SAO/NASA ADS)

¹ Τμήμα Φυσικής, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

² Εργαστήριο Τεχνητής Νοημοσύνης και Ανάλυσης Πληροφοριών, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

³ Τμήμα Αγγλικής Γλώσσας και Λογοτεχνίας, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Περιεχόμενα

Περίληψη	i
Ευχαριστίες	iii
Κατάλογος σχημάτων	ix
Κατάλογος πινάκων	xi
Ορολογία	xiii
1 Εισαγωγή	1
1.1 Ιδιότητες των αστρονομικών χρονοσειρών	1
1.2 Θόρυβος	3
1.3 Η περιοδικότητα στην αστρονομία	5
1.4 Αναζήτηση περιόδου	6
1.4.1 Αναδίπλωση	7
1.4.2 Fast Folding Algorithm	11
1.4.3 Ανάλυση Fourier	15
1.4.4 Radix-2 Fast Fourier Transform	21
1.5 Ιδιότητες συγκεκριμένων τύπων καμπυλών φωτός	25
1.5.1 Κηφείδες	25
1.5.2 Διπλοί εκλειπτικοί αστέρες	27
1.5.3 Διαβάσεις εξω-ηλιακών πλανητών	30
1.6 Στόχοι της εργασίας	32

2	Μεθοδολογία	35
2.1	Προγραμματισμός	35
2.1.1	Συμβάσεις	35
2.1.2	Πραγματοποίηση κώδικα FFT, FFA και αναδίπλωσης . .	36
2.2	Τεστ υπολογιστικής πολυπλοκότητας	40
2.3	Τεστ σε κατασκευασμένες καμπύλες φωτός	41
2.4	Τεστ σε δεδομένα της αποστολής Kepler	43
2.4.1	Επιλογή των στόχων	43
2.4.2	Διαχείριση καμπυλών φωτός	44
2.4.3	Τακτοποίηση χρονοσειρών σχεδόν ίσων διαστημάτων . .	44
2.4.4	Ανάλυση των δεδομένων	47
3	Αποτελέσματα	49
3.1	Τεστ υπολογιστικής πολυπλοκότητας	49
3.2	Τεστ σε κατασκευασμένες καμπύλες φωτός	53
3.3	Ανάλυση δεδομένων του Kepler	58
3.3.1	Παράδειγμα εξόδου προγράμματος	58
3.3.2	Αποτελέσματα	59
4	Συζήτηση	85
4.1	Τεστ υπολογιστικής πολυπλοκότητας	85
4.2	Τεστ σε κατασκευασμένες καμπύλες φωτός	86
4.3	Ανάλυση δεδομένων του Kepler	88
5	Συμπεράσματα	93
6	Προτάσεις	95
6.1	Αλλαγές	95
6.1.1	Συμβάσεις προγραμματισμού	95
6.1.2	Κατάλληλα στατιστικά μέτρα σημαντικότητας	96
6.1.3	Τακτοποίηση	96
6.2	Μελλοντικές προεκτάσεις	96
6.2.1	Μελέτη τάσεων και έγχρωμων θορύβων	96
6.2.2	Συνεργασία FFT/FFA/LS	97

Παράρτημα Α΄ Μαθηματικοί ορισμοί και αποδείξεις	99
Α΄.1 Big O notation	99
Α΄.2 Υπολογιστική πολυπλοκότητα για ‘τυφλές’ έρευνες με FFA	100
Α΄.3 Τύποι τακτοποίησης	101
Παράρτημα Β΄ Γεννήτριες τυχαίων αριθμών	103
Β΄.1 Ομοιόμορφη κατανομή	103
Β΄.2 Κανονική κατανομή	104
Παράρτημα Γ΄ Κύριος κώδικας	105
Γ΄.1 Τύποι μεταβλητών	105
Γ΄.2 Ρουτίνα τακτοποίησης	106
Γ΄.3 Κώδικας FFT	107
Γ΄.4 Κώδικας FFA	109
Γ΄.5 Ρουτίνα αναδίπλωσης	113
Γ΄.6 Ρουτίνα Re-binned αναδίπλωσης	114
Ενρετήριο	117
Βιβλιογραφία	119

Κατάλογος σχημάτων

1.1	Καμπύλη φωτός του HIP 97439	6
1.2	Αριθμοί Wolf των ηλιακών κηλίδων (1700–1988)	6
1.3	Πρώτο διάγραμμα ροής του Fast Folding Algorithm	11
1.4	Δεύτερο διάγραμμα ροής του Fast Folding Algorithm	12
1.5	Ανάλυση του FFA	16
1.6	Παράδειγμα ανάλυσης και σύνθεσης με DFT	19
1.7	Φασματική διάχυση στον DFT	20
1.8	FFT ‘πεταλούδα’	23
1.9	Διάγραμμα ροής του FFT	24
1.10	Ιδιότητες καμπυλών φωτός του δ Cephei	26
1.11	Περιγραφή των διπλών εκλειπτικών συστημάτων	27
1.12	Ιδιότητες καμπυλών φωτός διπλών εκλειπτικών	28
1.13	Τριδιάστατη αναπαράσταση μία διάβασης εξωπλανήτη	31
1.14	Διάφορα προφίλ διαβάσεων	32
2.1	Χρόνοι δειγματοληψίας στην αποστολή Kepler	44
2.2	Ασυνέχειες στη δειγματοληψία του Kepler	45
3.1	Αποτελέσματα ταχύτητας του FFA	51
3.2	Αποτελέσματα πολυπλοκότητας FFA	51
3.3	Αποτελέσματα ταχύτητας του FFT	52
3.4	Αποτελέσματα πολυπλοκότητας FFT	52
3.5	Ανίχνευση ημιτονοειδών σημάτων	53
3.6	Σφάλματα εντοπισμού ημιτονοειδών σημάτων	54

3.7	Ανίχνευση ορθογώνιων σημάτων κύκλου εργασίας 20%	55
3.8	Ανίχνευση ορθογώνιων σημάτων κύκλου εργασίας 10%	56
3.9	Ανίχνευση ορθογώνιων σημάτων κύκλου εργασίας 5%	57
3.10	Αποτελέσματα (a) για KID 7548061	60
3.11	Αποτελέσματα (b) για KID 7548061	61
3.12	Αποτελέσματα (a) για KID 7899428	62
3.13	Αποτελέσματα (b) για KID 7899428	63
3.14	Αποτελέσματα (a) για KID 11347875	64
3.15	Αποτελέσματα (b) για KID 11347875	65
3.16	Αποτελέσματα (a) για KID 12406908	66
3.17	Αποτελέσματα (b) για KID 12406908	67
3.18	Αποτελέσματα (a) για KID 4660997	68
3.19	Αποτελέσματα (b) για KID 4660997	69
3.20	Αποτελέσματα (a) για KID 4544587	70
3.21	Αποτελέσματα (b) για KID 4544587	71
3.22	Αποτελέσματα (a) για KID 8868650	72
3.23	Αποτελέσματα (b) για KID 8868650	73
3.24	Αποτελέσματα (a) για KID 10618251	74
3.25	Αποτελέσματα (b) για KID 10618251	75
3.26	Αποτελέσματα (a) για KID 11446443	76
3.27	Αποτελέσματα (b) για KID 11446443	77
3.28	Αποτελέσματα (a) για KID 10666592	78
3.29	Αποτελέσματα (b) για KID 10666592	79
3.30	Αποτελέσματα (a) για KID 9941662	80
3.31	Αποτελέσματα (b) για KID 9941662	81
3.32	Αποτελέσματα (a) για KID 11414511	82
3.33	Αποτελέσματα (b) για KID 11414511	83
4.1	Διορθωμένη καμπύλη φωτός για τον KID 4660997	91
4.2	Διορθωμένη καμπύλη φωτός για τον KID 4544587	91
4.3	Διορθωμένη καμπύλη φωτός για τον KID 8868650	92
4.4	Διορθωμένη καμπύλη φωτός για τον KID 10618251	92

Κατάλογος πινάκων

1.1	Παράδειγμα του πλήθους των αναδιπλώσεων που πραγματοποιούνται από τον FFA	14
1.2	Παράδειγμα αναστροφής bit	24
3.1	Αποτελέσματα τεστ υπολογιστικής πολυπλοκότητας	50
3.2	Σύνοψη αποτελεσμάτων στα δεδομένα του Kepler	59

Ορολογία

Όρος	Εξήγηση
ASAS	All Sky Automated Survey
ASCII	American Standard Code for Information Interchange κωδικοποίηση
BJD	Βαρυκεντρική Ιουλιανή μέρα
BKJD	Βαρυκεντρική Ιουλιανή μέρα της αποστολής Kepler. Ισούται με $BJD - 2454833.0$ (0 το μεσημέρι της 1-1-2009)
Dec	Απόκλιση
DFT	Διακριτός Μετασχηματισμός Fourier
ETS	Ισαπέχουσες χρονοσειρές (σύνολο μεγεθών, συνο- δευόμενο από δύο χρόνους: αρχή παρατήρησης και απόσταση μεταξύ παρατηρήσεων)
FFA	Fast Folding Algorithm
FFT	Fast Fourier Transform — σε αυτό το κείμενο αναφέ- ρεται στον αλγόριθμο Radix-2 Fast Fourier Transform
GTS	Γενική χρονοσειρά
J2000.0	Εποχή 1η Ιανουαρίου 2000, 12h TT (JD 2451545.0 TT). Σε αυτό το κείμενο αναφέρεται στο ουρανογρα- φικό σύστημα αναφοράς που ορίζεται από το μέσο ισημερινό και ισημερία του J2000.0

συνέχεια στην επόμενη σελίδα

συνέχεια από προηγούμενη σελίδα

JD	Η συνεχής απαρίθμηση των ημερών που έχουν πα- ρέλθει από το μέσο μεσημέρι του Greenwich της 1ης Ιανουαρίου του 4713 π.Χ.
KID	Kepler ID: προσδιοριστικό αστέρα στο πεδίο της απο- στολής Kepler
KOI	Kepler Object of Interest: Αντικείμενο ενδιαφέροντος Kepler
LSM	Μέθοδος ελαχίστων τετραγώνων
NASA	National Aeronautics and Space Administration: η δια- στημική υπηρεσία των Ηνωμένων Πολιτειών
SAO/NASA ADS	Smithsonian Astrophysical Observatory/NASA Astrophysics Data System: μία ψηφιακή βιβλιοθήκη επιστημονικών δημοσιεύσεων
RA	Ορθή αναφορά
SNR	Λόγο σήματος – θορύβου
SIMBAD	Set of Identifications, Measurements and Bibliography for Astronomical Data: μία αστρονομική βάση δεδομένων που λειτουργεί υπό το Centre de Données astronomiques de Strasbourg (CDS)
TT	Γήινος χρόνος
UST	Μη τακτική χρονοσειρά: σύνολο μεγεθών και χρόνων παρατήρησης

“ $2+2 \simeq 5$ ”

Sheuer, Peter (Γερμανός Αστρονόμος)

1

Εισαγωγή

1.1 Ιδιότητες των αστρονομικών χρονοσειρών

Σχεδόν όλη η πληροφορία που λαμβάνουμε από τα ουράνια σώματα, φτάνει σε μας υπό τη μορφή φωτονίων¹. Τα αστρονομικά δεδομένα είναι κατανομές ιδιοτήτων φωτονίων: θέση ($\vec{r}$), χρόνος (t), διεύθυνση (π.χ. σε ουρανογραφικές συντεταγμένες: α , δ), ενέργεια (E) και πόλωση (παράμετροι Stokes)(Horne, 2012). Για παράδειγμα, μία αστροφωτογραφία είναι η κατανομή της ενέργειας (=συχνότητα=χρώμα) και διεύθυνσης (θέση του κάθε εικονοστοιχείου στη φωτογραφία) των φωτονίων που ανιχνεύθηκαν από τη φωτογραφική διάταξη.

Σε πολλές περιπτώσεις, οι ανωτέρω ιδιότητες μελετώνται καθώς εξελίσσονται στο χρόνο. Για παράδειγμα μία ταινία είναι η εξέλιξη της κατανομής του προηγούμενου παραδείγματος και κάθε καρέ αντιστοιχεί σε έναν χρόνο παρατήρησης. Γενικά, ένα σύνολο παρατηρήσεων συλλεγόμενων σε διαδοχικά σημεία του χρό-

¹ σήμερα, υπάρχει επίσης αστρονομία νετρίνων, ενώ πολλά υποσχόμενη είναι η έρευνα για την παρατήρηση βαρυτικών κυμάτων

νου ονομάζεται **χρονοσειρά** (Cowpertwait & Metcalfe, Cowpertwait & Metcalfe).

Δε χρειάζεται να ειπωθούν πολλά για τη σημασία των αστρονομικών χρονοσειρών: από την αρχαιότητα η μέτρηση του χρόνου βασιζόταν στην παρατήρηση των αλλαγών της θέσης ουρανίων σωμάτων (Ήλιο, Σελήνη ή ακόμα και Αφροδίτη). Περαιτέρω, πολλοί λόγιοι είδαν στη μελέτη προτύπων σε αστρονομικές χρονοσειρές τη γέννεση του ορθολογισμού και της σύγχρονης επιστήμης (Zakai, 2009).

Τακτικές και μη τακτικές χρονοσειρές

Μετρήσεις που λαμβάνονται σε ίσα διαστήματα χρόνου (π.χ. κάθε ώρα ή κάθε 11η Απριλίου) παρουσιάζουν μία πιο καθαρή εικόνα του παρατηρούμενου αντικειμένου από ότι αν ανήκαν σε μία μη τακτική χρονοσειρά. Επίσης, πολλές μέθοδοι ανάλυσης απαιτούν τέτοια ομοιομορφία (π.χ. FFA and FFT που παρουσιάζονται στην εργασία αυτή). Έτσι οι χρονοσειρές θεωρούνται συνήθως τακτικές και αποτελούν στόχο της πλειονότητας των παρατηρήσεων και πειραμάτων πολλών επιστημονικών πεδίων.

Η σταθερή δειγματοληψία είναι αρκετά δύσκολο εγχείρημα για την παρατηρησιακή αστροφυσική. Στις περιπτώσεις επίγειων τηλεσκοπίων, οι χρόνοι παρατήρησης περιορίζονται από πληθώρα παραμέτρων: κλιματικές (π.χ. νέφωση), τροχιακές (π.χ. εναλλαγή μέρας – νύχτας) ή ακόμα και κοινωνικές (π.χ. μη προγραμματισμένη αεροναυτική δραστηριότητα). Συνεπώς, οι χρονοσειρές σταθερή δειγματοληψίας μπορούν να παραχθούν σε λίγες μόνο περιπτώσεις (π.χ. παρατηρήσεις pulsar σε ραδιοφωνικά μήκη κύματος, που δεν επηρεάζονται από την εναλλαγή μέρας – νύχτας ή των περισσότερων κλιματικών συνθηκών και που συχνά λεπτά ή ώρες παρατηρήσεις είναι αρκετός χρόνος).

Εν αντιθέσει, τα διαστημικά παρατηρητήρια, λόγω της απουσίας κλίματος (ωστόσο πρέπει να λαμβάνεται υπόψη η ηλιακή δραστηριότητα και η κοσμική ακτινοβολία) και της ελευθερίας καθορισμού των τροχιακών παραμέτρων, μπορούν να παράγουν μετρήσεις σε σταθερά χρονικά διαστήματα.

Μαθηματικός συμβολισμός

Θα χρησιμοποιήσουμε καλλιγραφικά κεφαλαία γράμματα ($\mathcal{A}, \mathcal{B}, \dots$) για να συμβολίσουμε τις χρονοσειρές ή συναρτήσεις που επιδρούν σ' αυτές.

- **Μη τακτική χρονοσειρά (USTS):** μία πεπερασμένη μαθηματική ακολουθία N διανυσμάτων, το καθένα από τα οποία αντιστοιχεί στο μέτρο (x_i) μίας συγκεκριμένης ποσότητας και σε έναν χρόνο παρατήρησης (t_i):

$$\mathcal{X} = (x_i, t_i)_{i=0}^{N-1} \quad (1.1)$$

όπου $t_{i+1} > t_i$.

- **Τακτική χρονοσειρά (ESTS):** το σύνολο τριών αντικειμένων: (i) μία πεπερασμένη μαθηματική ακολουθία N μέτρων (x_i) μίας ορισμένης ποσότητας, (ii) ο χρόνος αφετηρίας της παρατήρησης (τ_0) και (iii) η περίοδος δειγματοληψίας (p_s). Συνεπώς, η παρατήρηση (x_i) αντιστοιχεί σε χρόνο $\tau_i = \tau_0 + i \cdot p_s$.

$$\mathcal{Y} = \{(x_i)_{i=0}^{N-1}, \tau_0, p_s\} \quad (1.2)$$

- **Γενική χρονοσειρά (GTS):** Χάριν απλότητας, όταν δεν ενδιαφερόμαστε για την αφετηρία των παρατηρήσεων ή την περίοδο δειγματοληψίας, θα συμβολίζουμε τις τακτικές χρονοσειρές ως

$$\mathcal{Z} = (x_i)_{i=0}^{N-1} \quad (1.3)$$

1.2 Θόρυβος

Μία ενδογενής ιδιότητα κάθε μέτρησης είναι η αβεβαιότητα. Μία χρονοσειρά παρατηρήσεων ενός φυσικού φαινομένου μπορεί να θεωρηθεί ως ένα σήμα με θόρυβο. Θα χρησιμοποιήσουμε ένα παράδειγμα για να δείξουμε καλύτερα το γεγονός αυτό.

Για να εκτιμήσουμε το μέγεθος ενός αστέρα χρησιμοποιώντας ένα οπτικό τηλεσκόπιο, μπορούμε να μετρήσουμε την ποσότητα ενέργεια που καταφθάνει από μία ορισμένη στερεά γωνία που περιλαμβάνει τον στόχο, για ένα χρονικό διάστημα. Αυτή η ενέργεια δεν είναι ακόμα αντιπροσωπευτική του μεγέθους του αστέρα. Στο πεδίο παρατήρησης υπάρχουν και άλλα αντικείμενα: γήινη ατμόσφαιρα, μεσοπλανητικό και μεσοαστρικό υλικό που διαχέει φως, μη εντοπισμένα σώματα κτλ.. Επίσης, τα φωτόνια είναι διακριτές οντότητες οπότε εισάγουν κβαντικό θόρυβο και η ανίχνευσή τους έχει πιθανοκρατική φύση. Επιπλέον, πάντα ανιχνεύονται και θερμικά φωτόνια λόγω του επίγειου χαρακτήρα των ανιχνευτικών διατάξεων ή της παρουσίας ηλεκτρονικών κυκλωμάτων στα διαστημικά τηλεσκόπια.

Για να υπερβούμε αυτές τις δυσκολίες, εφαρμόζεται ένας αριθμός τεχνικών επεξεργασίας δεδομένων για την μετατροπή των μετρήσεων σε χρήσιμη πληροφορία. Ακόμα όμως και τότε, οι αριθμητικές μέθοδοι που χρησιμοποιούνται και η φύση των ηλεκτρονικών υπολογιστών, εισάγουν θόρυβο λόγω της περιορισμένης ακρίβειας των αριθμητικών πράξεων.

Κανονικότητα

Η κατανομή της ‘στατιστικής καταμετρήσεων’, όπως στην περίπτωση της ανίχνευσης φωτονίων στην αστρονομία, είναι η κατανομή Poisson. Ωστόσο η επιρροή από μία πληθώρα φυσικών διεργασιών και η συλλογή μεγάλου πλήθους δεδομένων, μας επιτρέπουν να υποθέσουμε ότι η κανονική (Γκαουσιανή) κατανομή είναι κατάλληλη – εξαιτίας του ‘κεντρικού οριακού θεωρήματος’ (Wall & Jenkins, 2003).

Συμβολισμός

Συνοψίζοντας, συμπληρωματικά στο συμβολισμό των παραγράφων Εξίσωση (1.1) και Εξίσωση (1.2), η ακολουθία των μετρήσεων (x_i) θα θεωρείται ως το άθροισμα ενός ψηφιακού σήματος (s_i) και θορύβου (n_i) κανονικής κατανομής:

$$x_i = s_i + n_i \quad (1.4)$$

Ένα μέτρο της ποιότητα μίας χρονοσειράς, είναι ο λόγος σήματος θορύβου

(SNR). Είναι ο λόγος της ισχύος ενός σήματος προς την ισχύ του θορύβου. Η ισχύς $\mathcal{P}$ μία χρονοσειράς (όπως ενός διακριτού σήματος) ορίζεται ως το άθροισμα των τετραγώνων των στοιχείων της (Sabin, 2008):

$$\mathcal{P}(\mathcal{Z}) = \sum_{i=0}^{N-1} x_i^2 \quad (1.5)$$

$$SNR(\mathcal{Z}) = \frac{\sum_{i=0}^{N-1} s_i^2}{\sum_{i=0}^{N-1} n_i^2} \quad (1.6)$$

Το μεγαλύτερο εμπόδιο στον προσδιορισμό του λόγου σήματος – θορύβου είναι το γεγονός ότι σε μία παρατήρηση (σε αντίθεση με ένα υπολογιστικό μοντέλο) δε μπορούμε να ξεχωρίσουμε το σήμα από το θόρυβο. Υπάρχουν μέθοδοι (στατιστικές, από τη θεωρία πληροφορίας, κτλ.) για την αντιμετώπιση του προβλήματος αλλά εξαρτώνται πάντα από ερμηνείες και εκτιμήσεις για τα δεδομένα.

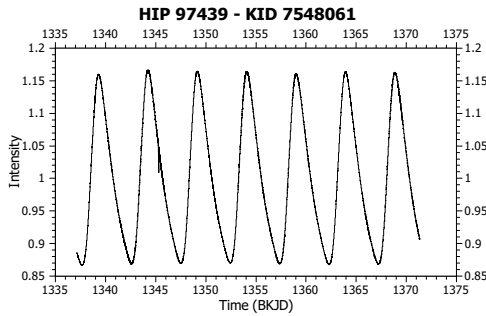
1.3 Η περιοδικότητα στην αστρονομία

Ένα από τα πιο σημαντικά ευρήματα σε μία χρονοσειρά είναι η περιοδικότητα, υπό την έννοια ότι επιτρέπει προβλέψεις και προάγει την ανακάλυψη υποβοσκομένων φυσικών διεργασιών. Μερικά παραδείγματα περιοδικών αστρονομικών ποσοτήτων ή φαινομένων είναι:

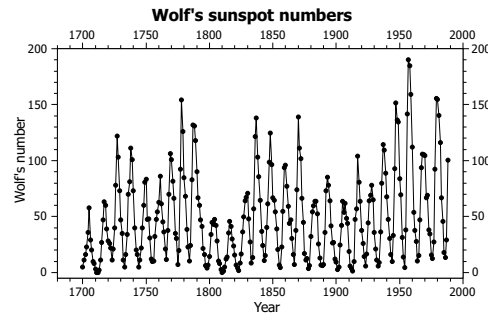
- το φως από παλλόμενους αστέρες, περιστρεφόμενους μη σφαιρικούς αστέρες, διπλά εκλειπτικά συστήματα ¹ (Σχήμα 1.1)
- διαβάσεις εξωηλιακών πλανητών
- ο 11-ετής (ή 22-ετής) κύκλος μαγνητικής δραστηριότητας του Ήλιου (Σχήμα 1.2)

¹ σε αντίθεση με τους εκρηκτικούς μεταβλητούς όπως οι καινοφανείς, των οποίων η μεταβλητότητα δεν ακολουθεί ένα επαναλαμβανόμενο μοτίβο

- η ακτινική ταχύτητα διπλών συστημάτων ή αστερών με εξωπλανήτες μεγάλης μάζας
- η ακτινοβολία των pulsar στα ραδιοκύματα



Σχήμα 1.1: Καμπύλη φωτός του HIP 97439 (ή KIC 7548061). Πρόκειται για μία κλασική περίπτωση Κηφείδη μεταβλητού με προφανή περιοδικότητα (Thomson et al., 2013).



Σχήμα 1.2: Οι αριθμοί Wolf των ηλιακών κηλίδων υποδεικνύουν περιοδικότητα στην ηλιακή μαγνητική δραστηριότητα (Hipel & McLeod, 1994)

Φυσικά, διαφορετικού τύπου μεταβολές — περιοδικές ή μη — μπορούν να συνυπάρχουν σε ένα υποκείμενο παρατήρησης. Παράδειγμα αυτού είναι ο μεταβλητός αστέρας τύπου δ Scuti, WASP-33 (ή HD 15082) στην καμπύλη φωτός του οποίου έχουν ανιχνευτεί διαβάσεις εξωπλανήτη (Herrero et al., 2011).

1.4 Αναζήτηση περιόδου

Η περιοδικότητα σε μία καμπύλη φωτός όπως αυτής του Σχήμα 1.1 μπορεί να μοιάζει προφανής ακόμα και με γυμνό οφθαλμό αλλά αυτό δεν είναι ο κανόνας:

- υψηλά επίπεδα θορύβου μπορούν να κρύψουν την περιοδικότητα (και άλλες τάσεις), ειδικά όταν παρατηρούμε μακρινά αντικείμενα
- η διακριτή φύση των χρονοσειρών μπορεί να κρύψει τα σχήματα σε μία καμπύλη φωτός όταν η περίοδος δειγματοληψίας είναι συγκρίσιμη με την περίοδο του ζητούμενου φαινομένου

- η παρουσία πολλών διαφορετικών περιόδων μπορεί να ‘ανακατέψει’ την κυματομορφή δημιουργώντας την εντύπωση μίας τυχαίας κατανομής
- χρειαζόμαστε στατιστικά μεγέθη προσδιορισμού της περιοδικότητας που να μπορούν να εφαρμοστούν αυτόματα στους αναρίθμητους στόχους των σύγχρονων ερευνών (μερικές φορές εκατομμύρια)

Υπάρχουν πολλές προσεγγίσεις στο πρόβλημα αλλά εδώ θα ασχοληθούμε με δύο μεθόδους (αυτό)συσχέτισης και¹: (i) αναδίπλωση και (i) διακριτός μετασχηματισμός Fourier.

1.4.1 Αναδίπλωση

Αν μία χρονοσειρά περιγράφει ένα περιοδικό φαινόμενο τότε οι περίοδοι θα τείνουν να είναι παραπλήσιες, αλλά όχι ακριβώς ίδιες για τους λόγους που εξηγήσαμε παραπάνω (θόρυβος, άλλες συνιστώσες, κτλ.). Η αυτοσυσχέτιση των δεδομένων με την υπέρθεση πολλών περιόδων θα οδηγούσε σε μία **αναδιπλωμένη χρονοσειρά ή φασικό διάγραμμα** διάρκειας όση και η περίοδος. Το αναδιπλωμένο προφίλ θα αποτελεί (όπως θα δείξουμε αργότερα) μία ενίσχυση του σήματος σε σχέση με το θόρυβο.

Βελτίωση λόγου σήματος – θορύβου

Ας ρίξουμε μία ματιά σε μία απλή περίπτωση αναδίπλωσης: δύο ίσου μεγέθους ακολουθίες (Z' και Z'') που αποτελούν δύο ολόκληρες περιόδους της αρχικής χρονοσειράς, αθροίζονται ώστε να δώσουν μία αναδιπλωμένη χρονοσειρά (Z). Ιδανικά, τα σήματα είναι ίδια ($s'_i = s''_i$), αλλά καθώς ο θόρυβος θεωρείται τυχαίος και ανεξάρτητος του σήματος (μηδενικός συντελεστής αυτοσυσχέτισης Pearson) θα διαφέρει στις δύο περιπτώσεις και η άθροισή του δε θα τον διπλασιάσει. Πιο συγκεκριμένα,

¹ η διαδικασία σύγκρισης μίας χρονοσειράς με τον εαυτό της (με διαφορά στο χρόνο) ή μία συνάρτηση ή μία άλλη χρονοσειρά

$$\begin{aligned}
\mathcal{Z} &= \mathcal{Z}_1 + \mathcal{Z}_2 \\
x_i &= x'_i + x''_i \\
s_i + n_i &= s'_i + n'_i + s''_i + n''_i \\
s_i &\approx 2s'_i \text{ και } n_i \approx n'_i\sqrt{2}
\end{aligned} \tag{1.7}$$

Στην πραγματικότητα, ο θόρυβος n_i δε μπορεί να είναι ανάλογος του θορύβου σε οποιαδήποτε από τις δύο χρονοσειρές. Η έκφραση $\sqrt{2}$ δηλώνει το γεγονός ότι η πρόσθεση n τυχαίων μεταβλητών Γκαουσιανής κατανομής με την ίδια τυπική απόκλιση, οδηγεί σε τυχαίους αριθμούς της ίδια κατανομής αλλά με τυπική απόκλιση ενισχυμένη κατά $\sqrt{n}$, μία συνέπεια του ‘νόμου των μεγάλων αριθμών’ (Spiegel, 1975).

Αν η περίοδος περιέχεται στα δεδομένα m φορές, η συνολική βελτίωση του SNR θα είναι:

$$\frac{SNR_{fold}}{SNR_{data}} = \sqrt{m} \tag{1.8}$$

Εκτίμηση περιόδου

Η περίοδος του σήματος είναι στην πράξη άγνωστη και μπορούμε μόνο να εφαρμόσουμε την παραπάνω διαδικασία (§ 1.4.1) για δοκιμαστικές περιόδους. Η μέθοδος θα δώσει μέγιστη ενίσχυση για την περίπτωση όπου η δοκιμαστική περίοδος είναι πολύ κοντά σ’ αυτήν του φαινομένου. Η επανάληψη για ένα μεγάλο εύρος δοκιμαστικών περιόδων μπορεί να οδηγήσει σε μία καλή εκτίμηση της περιόδου. Το παρακάτω είναι ένα απλουστευμένο παράδειγμα:

Φανταστείτε έναν άνθρωπο που επαναλαμβάνει τους 8 πρώτους φυσικούς αριθμούς. Οι παύσεις μεταξύ των αριθμών είναι 10 δευτερόλεπτα. Δηλαδή, η περίοδος της χρονοσειράς είναι 80 δευτερόλεπτα.

Αν ένα δεύτερο άτομο μπαίνει και βγαίνει (μετά από 5 δευτερόλεπτα) από το δωμάτιο στο οποίο βρίσκεται το πρώτο, κάθε 50 δευτερόλεπτα, θα καταγράψει την επανάληψη της ακολουθίας : “πέντε, δύο, εφτά, τέσσερα, ένα, έξι, τρία, οχτώ”¹. Οι αριθμοί αυτοί μοιάζουν τυχαίοι στον παρατηρητή και το άθροισμά

¹ Η ισότητα των μεγεθών των δύο χρονοσειρών και η εμφάνιση ενός ‘τέλειου’ ανακατέματος

τους θα είναι 36. Αν αργούσε λίγο στην αρχή (φάση), θα κατέληγε με την ίδια εικόνα.

Τώρα, ένας άλλος παρατηρητής, μπαίνει και βγαίνει από το δωμάτιο κάθε 80 δευτερόλεπτα (δηλαδή όσο και οι περίοδος του σήματος του εκφωνητή των αριθμών). Ανάλογα με το πότε αποφασίσει να αρχίσει την ‘εισβολή’ του για πρώτη φορά στο δωμάτιο, θα καταγράψει “ένα, ένα, ένα, ...” ή “εφτά, εφτά, ...” κτλ. Αμέσως θα αντιληφθεί ένα αξιοσημείωτο μοτίβο και θα σημειώσει ένα άθροισμα που θα είναι μικρό ή μεγάλο (σε σχέση με αυτό του πρώτου προσώπου) ανάλογα τη φάση του.

Η αναδίπλωση είναι η ίδια διαδικασία: εισερχόμαστε σε ένα δωμάτιο με διαφορετικές περιόδους με μία ομάδα παρατηρητών για διάφορες φάσεις. Σε κάθε ομάδα, ένα παρατηρητής καταγράφει μεγάλες τιμές για το άθροισμα — κάτι που αντιστοιχεί σε υψηλή πιθανότητα να έχει ανήκει στην ομάδα που έχει τη σωστή περίοδο. Μία απλή αναφορά των ομάδων θα υποδείκνυε τη σωστή περίοδο του σήματος.

Σύνοψη της μεθόδου

1. Ως είσοδο έχουμε τη χρονοσειρά

$$\mathcal{Z} = (x_i), \text{ όπου } i \in \{0, 1, \dots, N - 1\} \quad (1.9)$$

2. Καθορίζεται μία αρχική δοκιμαστική περίοδος P

3. Αναδιπλώνουμε τα δεδομένα σε μία χρονοσειρά αυτής της διάρκειας

$$\mathcal{F} = (f_j), \text{ όπου } j \in \{0, 1, \dots, P - 1\} \quad (1.10)$$

4. Κάθε στοιχείο του φασικού διαγράμματος (phase bin) f_j υπολογίζεται από τον τύπο

$$f_j = \frac{\sum_{k=0}^{m_j-1} x_{j+kP}}{m_j} \quad (1.11)$$

είναι απόρροια του γεγονότος ότι οι αριθμοί 5 και 8 είναι πρώτοι μεταξύ τους

όπου m_j είναι το πλήθος των αναδιπλώσεων. Η κανονικοποίηση με αυτήν τη διαίρεση είναι απαραίτητη γιατί η δοκιμαστική περίοδος P δε διαιρεί πάντα το μέγεθος της εισόδου N . Έτσι, η τιμή του m_j δεν είναι ίδια για όλα τα phase bins (j) καθώς μπορούμε να αναδιπλώσουμε, e.g. 8 αριθμούς σε μία χρονοσειρά 3 στοιχείων: το πρώτο και το δεύτερο θα έχουν το μέσο όρο των τριών μεγεθών ($m_j=3$) και το τρίτο, δύο μεγεθών ($m_j=2$). Συνεπώς,

$$m_j = \begin{cases} \lfloor \frac{N}{P} \rfloor + 1 & \text{για } i < N \pmod{P} \\ \lfloor \frac{N}{P} \rfloor & \text{αλλιώς} \end{cases} \quad (1.12)$$

5. Υπολογίζεται η ισχύς της αναδιπλωμένης χρονοσειράς
6. Επιλέγουμε μία νέα δοκιμαστική περίοδο (μέχρι κάποιο μέγιστο) και επιστρέφουμε στο βήμα 2
7. Αφού έχουν δοκιμαστεί όλες οι περίοδοι, ένα διάγραμμα ισχύος – περιόδου μπορεί να κατασκευαστεί
8. Αν ενδιαφερόμαστε για το το πιο δυνατό περιοδικό σήμα, η περιόδός του θα είναι αυτή του μέγιστου του προηγούμενου διαγράμματος (peak period)

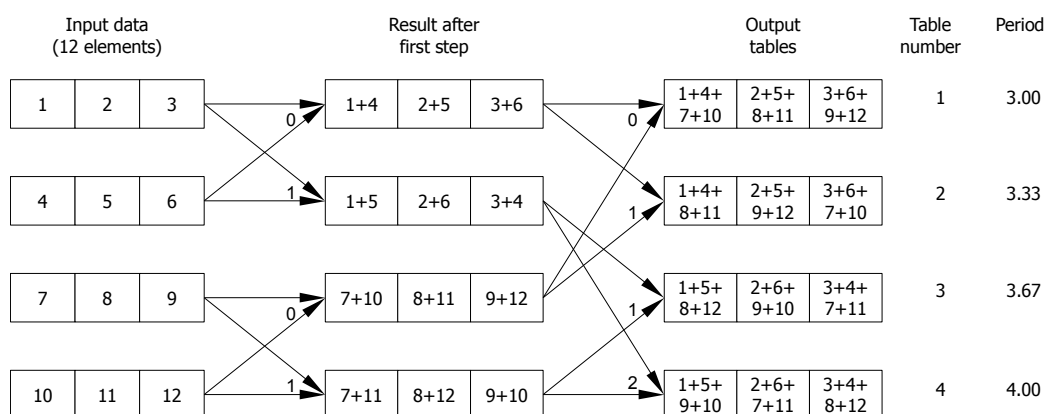
Σε μία brute-force πραγματοποίηση του αλγορίθμου, αυτός έχει υπολογιστική πολυπλοκότητα (βλ. Παράρτημα Α΄.1) ίση με $O(n^2)$. Το πλήθος των αριθμητικών πράξεων που απαιτούνται για κάθε δοκιμαστική περίοδο είναι ανάλογη του N : (i) πρόσθεση N αριθμών κατά ομάδες, (ii) N αναγνώσεις για τον υπολογισμό της ισχύος της αναδίπλωσης. Σε ‘τυφλές’ έρευνες περιόδου, όπου ελέγχουμε για περιόδους σε ένα διάστημα που εξαρτάται από το μέγεθος της εισόδου (π.χ. από δύο στοιχεία ως $N/2$ στοιχεία), οπότε η συνολική πολυπλοκότητα index Υπολογιστική πολυπλοκότητα είναι $O(n^2)$

Η τεχνική της αναδίπλωσης εφαρμόζεται επίσης για να αναδείξει το σχήμα του περιοδικού σήματος ή όπως λέγεται ‘προφίλ’. Πραγματοποιώντας αναδίπλωση χρησιμοποιώντας μία εκτίμηση της περιόδου ενισχύεται το σήμα, δίνεται ένας μέσος όρος (όχι πάντα επιθυμητό: υπάρχουν σχεδόν περιοδικές φυσικές διεργασίες) πολλών περιόδων και μικραίνει ο όγκος της πληροφορίας, επιταχύνοντας άλλες επεξεργασίες.

1.4.2 Fast Folding Algorithm

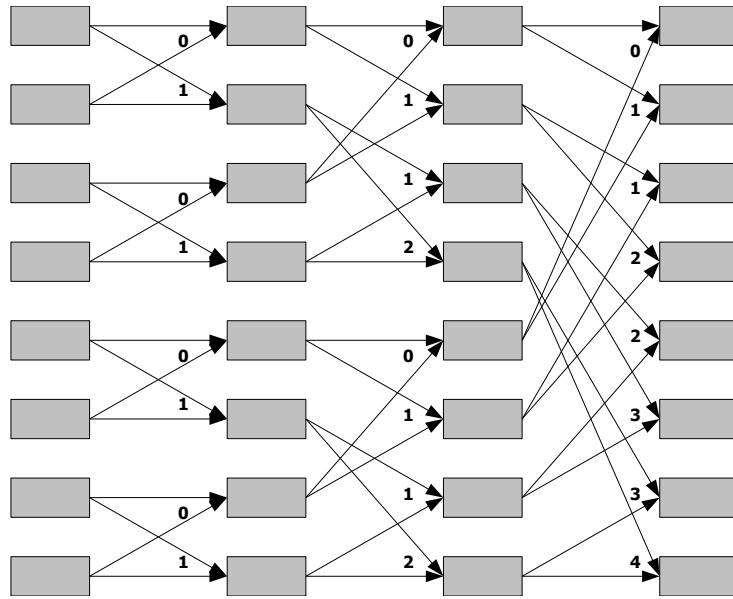
Η παραπάνω brute-force προσέγγιση μπορεί να βελτιστοποιηθεί αρκετά. Ένας μεγάλος αριθμός προσθέσεων για την παραγωγή phase bin (Εξίσωση (1.11)) επαναλαμβάνεται πολλές φορές. Εμπνεόμενος από τον FFT των Cooley και Tukey (Cooley & Tukey, 1965), τέσσερα χρόνια αργότερα, ο David Staelin δημιούργησε έναν παρόμοιο αλγόριθμο, τον Fast Folding Algorithm (Staelin, 1969), για την πραγματοποίηση αναδιπλώσεων χρονοσειρών αποφεύγοντας περιττές αριθμητικές πράξεις.

Διάγραμμα ροής



Σχήμα 1.3: Διάγραμμα ροής του FFA όπως εφαρμόζεται σε μία χρονοσειρά 12 στοιχείων. Σημειώσεις: **(i)**. Οι αριθμοί στα κελιά υποδηλώνουν τη θέση των στοιχείων στην αρχική διάταξη, ενώ το σύμβολο της άθροισης (+) υποδηλώνει άθροιση των περιεχομένων των κελιών αυτών και όχι των θέσεων τους. **(ii)**. Τα βέλη δείχνουν τη ροή της πληροφορίας σε κάθε βήμα. Οι ακέραιοι αριθμοί ανάμεσα σε κάθε ζεύγος βελών, υποδεικνύει τον αριθμό των θέσεων κατά τον οποίο μεταπίπτουν δεξιά τα στοιχεία του υποπίνακα προέλευσης του δεύτερου βέλους. **iii**. Αυτή είναι μία δική μας αναπαραγωγή του πρώτου σχήματος στο Staelin (1969).

Η λειτουργία του FFA μπορεί να διαλευκανθεί με ένα σύντομο παράδειγμα. Μία χρονοσειρά 12 στοιχείων είναι να ελεγχθεί για σήμα στο εύρος περιόδων από 3 ως 4 φορές την περίοδο δειγματοληψίας — εν συντομία, ‘για περιόδους από 3 ως 4 στοιχεία’.



Σχήμα 1.4: Διάγραμμα ροής του FFA όπως αυτός εφαρμόζεται σε N στοιχεία για βασικής περιόδου $P = N/8$

Πρώτα, θα πρέπει να αναδιπλώσουμε τη χρονοσειρά στις περιόδους αυτές. Το Σχήμα 1.3 δείχνει τη μετατροπή της εισόδου σε τέσσερις αναδιπλώσεις, η καθεμία από τις οποίες αντιστοιχεί σε διαφορετική περίοδο στο επιθυμητό εύρος.

Ο αλγόριθμος αθροίζει τα στοιχεία ζευγαριών υποδιαιρέσεων του αρχικού πίνακα στοιχείων (ή του προηγούμενου βήματος) αφού μετακινήσει δεξιά τα στοιχεία ενός από αυτούς, κατά έναν παράγοντα. Το Σχήμα 1.4 δείχνει το διάγραμμα ροής στην περίπτωση 8 υποπινάκων ώστε ο αναγνώστης να αντιληφθεί καλύτερα την επιλογή ζευγαριών και μετακινήσεων¹.

Εφαρμογή του FFA σε n -δεδομένα για βασική περίοδο P_0

1. Ομαδοποίηση των αρχικών n δεδομένων σε M πίνακες P_0 στοιχείων ο καθένας. Καθώς οι υποπίνακες θα συνδυάζονται κατά ζεύγη σε συνεχώς διπλασιαζόμενες ομάδες πινάκων, ο M πρέπει να είναι δύναμη του 2, οπότε επιλέγουμε την μεγαλύτερη δυνατή τιμή προσπαθώντας να απορρίψουμε

¹ Οι μετακινήσεις υποδηλώνονται ανάμεσα στα βέλη, κάτι που δεν πρέπει να συγχέεται με την πράξη της ενίσχυσης στα κλασικά διαγράμματα ροής της θεωρίας επεξεργασίας σήματος

μόνο μέρος των αρχικών δεδομένων:

$$M = 2^{\left\lfloor \log_2 \frac{n}{P_0} \right\rfloor} \quad (1.13)$$

2. Ο συνολικός αριθμός στοιχείων, N , που θα χρησιμοποιηθεί θα είναι

$$N = M \cdot P_0 \leq n \quad (1.14)$$

3. Το πλήθος των βημάτων, S , είναι

$$S = \log_2 M = \left\lfloor \log_2 \frac{n}{P_0} \right\rfloor \quad (1.15)$$

4. M στο πλήθος, υποπίνακες (αναδιπλώσεις στο τελευταίο βήμα) θα σχηματιστούν: $\mathcal{F}_i$, όπου $i \in \{1, 2, \dots, M\}$ που αντιστοιχούν σε περίοδο $P(i)$:

$$P(i) = P_0 + \frac{i-1}{M-1} \quad (1.16)$$

5. Τώρα, οι αναδιπλώσεις είναι έτοιμες να κανονικοποιηθούν (π.χ. με διαίρεση δια του πλήθους των αθροίσεων, να αναλυθούν ή να εξεταστούν για μέγιστο στην ισχύ με σκοπό να εκτιμηθεί η περίοδος του σήματος όπως περιγράφεται στο § 1.4.1

Πολυπλοκότητα

Σε κάθε βήματα πραγματοποιούνται N προσθέσει, M μετακινήσεις, etc. Η πολυπλοκότητα του βήματος είναι $O(N)$ (βλ. Παράρτημα § Α'.1). Το πλήθος των βημάτων είναι $\log_2 N / P$, από το οποίο προκύπτει συνολική πολυπλοκότητα $O(N \log N)$.

Πολυπλοκότητα για 'τυφλες' έρευνες

Για να αναζητήσουμε περιοδικότητες σε ένα εύρος τιμών της περιόδου που εκτείνεται περισσότερο από όσο η περίοδος δειγματοληψίας, ο FFA πρέπει να

εκτελεστεί τόσος φορές όσες η μικρότερη διαφορά ανάμεσα σε ακεραίους που εμπεριέχουν το επιθυμητό εύρος. Παραδείγματος χάριν, για αναζήτηση στο εύρος (5.67, 9.21) ο αλγόριθμος πρέπει να ‘τρέξει’ 5 φορές — για βασικές περιόδους 5, 6, 7, 8, 9 και 10. Η συγκέντρωση των αναφορών κάθε εκτέλεσης μπορεί να οδηγήσει σε μία εκτίμηση της περιόδου του σήματος (π.χ. peak period.)

Η υπολογιστική πολυπλοκότητα για την εφαρμογή του FFA σε ένα μεγάλο εύρος περιόδων αποδεικνύεται ότι είναι ανάλογη του τετραγώνου της εισόδου ($O(n^2)$) — το ίδιο με την brute-force αναδίπλωση (βλ. § 1.4.1). Θα μπορούσε κανείς να αναρωτηθεί γιατί ο FFA είναι αποτελεσματικότερος.

Η απάντηση είναι κρυμμένη στην ανάλυση των δύο μεθόδων. Ο FFA αναδιπλώνει την αρχική χρονοσειρά όχι μόνο σε ακέραιες περιόδους αλλά και σε ένα πλήθος κλασματικών τιμών ανάμεσά τους. Για παράδειγμα, για $N=20$, και περιορίζοντας την αναζήτηση στο εύρος [2, 6]:

- **Brute-force:** 9 αναδιπλώσεις για δοκιμαστικές περιόδους 2, 3, 4, 5, 6, 7, 8, 9, 10
- **FFA:** ο αλγόριθμος θα εκτελεστεί για 9 περιόδους, παράγοντας 30 αναδιπλώσεις (Πίνακας 1.1)

Πίνακας 1.1: Παράδειγμα του πλήθους των αναδιπλώσεων που πραγματοποιούνται από τον FFA

Βασική περίοδος	Περίοδοι των αναδιπλώσεων
2	2.00, 2.14, 2.29, 2.43, 2.57, 2.71, 2.86, 3.00
3	3.00, 3.33, 3.66, 4.00
4	4.00, 4.33, 4.66, 5.00
5	5.00, 5.33, 5.66, 6.00
6	6.00, 7.00
7	7.00, 8.00
8	8.00, 9.00
9	9.00, 10.00
10	10.00, 11.00

Εδώ πρέπει να κάνουμε κάποιες σημειώσεις. Στον Πίνακας 1.1 μπορούμε να δούμε ότι οι ακέραιες περίοδοι αναδίπλωσης εμφανίζονται δύο φορές (εκτός από

τις οριακές τιμές 2.00 και 11.00). Έτσι, μόνο 22 διαφορετικές περίοδοι εκτιμώνται. Η εμπειρία μας έδειξα ότι η τελευταία αναδίπλωση κάθε εκτέλεσης του FFA είναι μία προσέγγιση (χαμηλότερη ισχύς) της πρώτης αναδίπλωσης μία άλλης εκτέλεσης για την ίδια περίοδο (βλ. § 1.4.1), με την τελευταία να είναι ίδια με την κλασική μέθοδο αναδίπλωσης. Το γεγονός αυτό μας αναγκάζει στην συμπερίληψη (i) όλων των διπλών εκτιμήσεων στην στατιστική και (ii) της τελευταίας εκτέλεσης του FFA (10.00, 11.00) — παρ' όλο που το εύρος καλύφθηκε από την προτελευταία εκτέλεση (9.00 and 10.00).

Αυτό σημαίνει πως η εκτίμηση της περιόδου με τον FFA είναι πολύ πιο ακριβής. Επίσης, η βελτιωμένη ανάλυση μπορεί να χρησιμοποιηθεί με το re-binning των δεδομένων: το μέγεθος της εισόδου μπορεί να μειωθεί και η συμπεριφορά $O(n^2)$ θα μειώσει έντονα το χρόνο εκτέλεσης σε σχέση με τον brute-force τρόπο, με την ανάλυση να παραμένει σε υψηλά επίπεδα.

Πλήθος αναδιπλώσεων και ανάλυση

Το συνολικό πλήθος αναδιπλώσεων, $C_{FFA}(N)$ εξαρτάται από το μέγεθος της εισόδου N και έχει πολύπλοκη έκφραση, γιατί η απαίτηση ο FFA να εκτελείται για είσοδο διατμημένη σε ομάδες πλήθους που είναι δύναμη του 2, εμπλέκει συναρτήσεις της θεωρίας αριθμών.

Αντί μίας αυστηρής μαθηματικής απόδειξης επιβεβαιώσαμε τις παρακάτω εξισώσεις, με υπολογιστικά πειράματα:

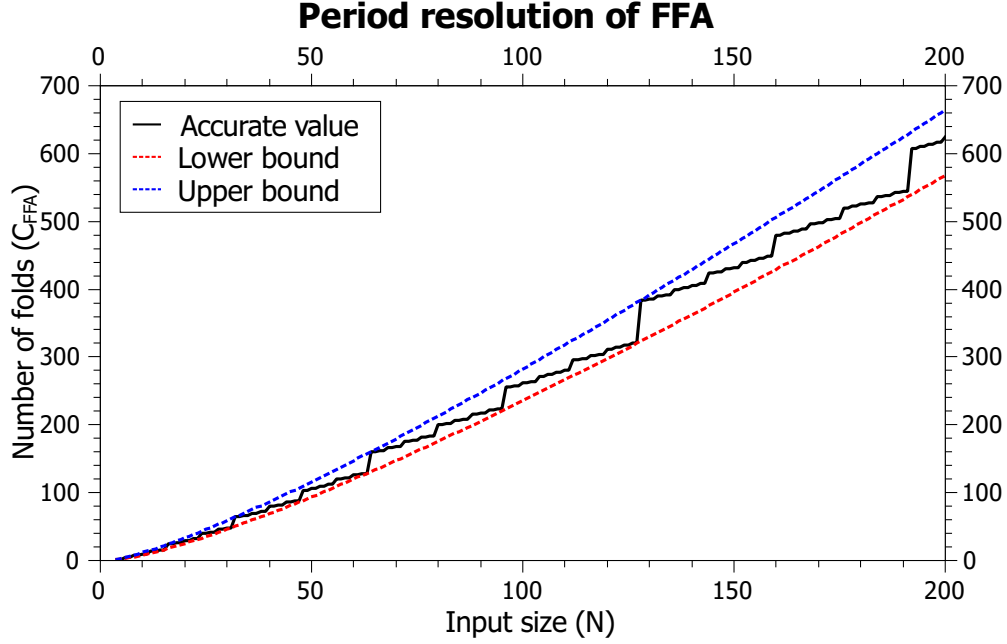
$$C_{FFA} \simeq \frac{1}{2}N \log_2 N - \frac{3}{4}N \quad (1.17)$$

Η συνάρτηση περιορίζεται πάντα από δύο άλλες (Σχήμα 1.5):

$$\frac{N+1}{2} \log_2 \frac{N+1}{4} \leq C_{FFA} \leq \frac{N}{2} \log_2 \frac{N}{2} \quad (1.18)$$

1.4.3 Ανάλυση Fourier

Μία χρονοσειρά είναι η απεικόνιση των δεδομένων στο χώρο του χρόνου. Είναι δυνατόν να αναπαραστήσουμε τα δεδομένα στο χώρο των συχνοτήτων με τους



Σχήμα 1.5: Η τιμή του C_{FFA} , (πλήθους αναδιπλώσεων και επομένως, πλήθους περιόδων που ελέγχονται), συναρτήσει του μεγέθους της εισόδου. Σημειώστε ότι οι διπλές εμφανίσεις συμπεριλαμβάνονται στη συνάρτηση (βλ. ‘αποτελεσματικότητα’).

μετασχηματισμούς Fourier. Εδώ θα τονιστούν μερικές βασικές αρχές της Ανάλυσης Fourier με σκοπό να καταλάβουμε τη λειτουργία του FFT (Chu, 2008).

Σειρά Fourier

Μία περιοδική συνάρτηση $x(t)$ μπορεί να εκφραστεί ως άθροισμα τριγωνομετρικών σειρών:

$$x(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{\pi n}{L}t\right) + b_n \sin\left(\frac{\pi n}{L}t\right) \right] \quad (1.19)$$

όπου

$$a_n = \frac{1}{L} \int_{-L}^L x(t) \cos\left(\frac{\pi n}{L}t\right) dt \quad (1.20)$$

$$b_n = \frac{1}{L} \int_{-L}^L x(t) \sin\left(\frac{\pi n}{L}t\right) dt \quad (1.21)$$

Γενίκευση για μιγαδικές συναρτήσεις

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{i\frac{\pi n}{L}t} \quad (1.22)$$

όπου

$$c_n = \frac{1}{2L} \int_{-L}^L x(t) e^{-i\frac{\pi n}{L}t} dt \quad (1.23)$$

Συνεχής μετασχηματισμός Fourier

Ο Συνεχής μετασχηματισμός Fourier (CFT) είναι η γενίκευση της σειράς Fourier που επιτρέπει την αναπαράσταση συνεχών μη περιοδικών σημάτων:

$$x(t) = \int_{-\infty}^{\infty} F(f) e^{-2\pi i f t} df \quad (1.24)$$

όπου

$$F(f) = \int_{-\infty}^{\infty} x(t) e^{2\pi i f t} dt \quad (1.25)$$

Διακριτός μετασχηματισμός Fourier

Για διακριτά σήματα σταθερής δειγματοληψίας (ή τακτικές χρονοσειρές) μεγέθους N , x_n όπου $n \in \{0, 1, \dots, N-1\}$, ο διακριτός μετασχηματισμός Fourier

(DFT), X_p όπου $p \in \{0, 1, \dots, N-1\}$, υπολογίζεται ως εξής:

$$X_p = \sum_{n=0}^{N-1} x_n e^{-i \frac{2\pi}{N} np} \quad (1.26)$$

Ο αντίστροφος μετασχηματισμός DFT μπορεί επίσης να υπολογιστεί:

$$x_n = \frac{1}{N} \sum_{p=0}^{N-1} X_p e^{i \frac{2\pi}{N} np} \quad (1.27)$$

Έτσι ένας αλγόριθμος που πραγματοποιεί DFT θα πρόσθετε και θα πολλαπλασιάσει N τιμές, N φορές, οδηγώντας σε υπολογιστική πολυπλοκότητα $O(N^2)$.

DFT: Φάσμα πλάτους και φάσης

Η έξοδος του DFT είναι μία σειρά μιγαδικών αριθμών. Το πρώτο στοιχείο είναι η σταθερά a_0 , δηλαδή η DC συνιστώσα του αρχικού σήματος. Τα επόμενα στοιχεία φέρουν την πληροφορία των αρμονικών. Το πλάτος των τελευταίων είναι $(|X_i|)$ το μέτρο των μιγαδικών αριθμών ενώ η φάση δίνεται από το όρισμα $\left(\tan^{-1} \frac{\text{Im}(X_i)}{\text{Re}(X_i)}\right)$ (Sabin, 2008).

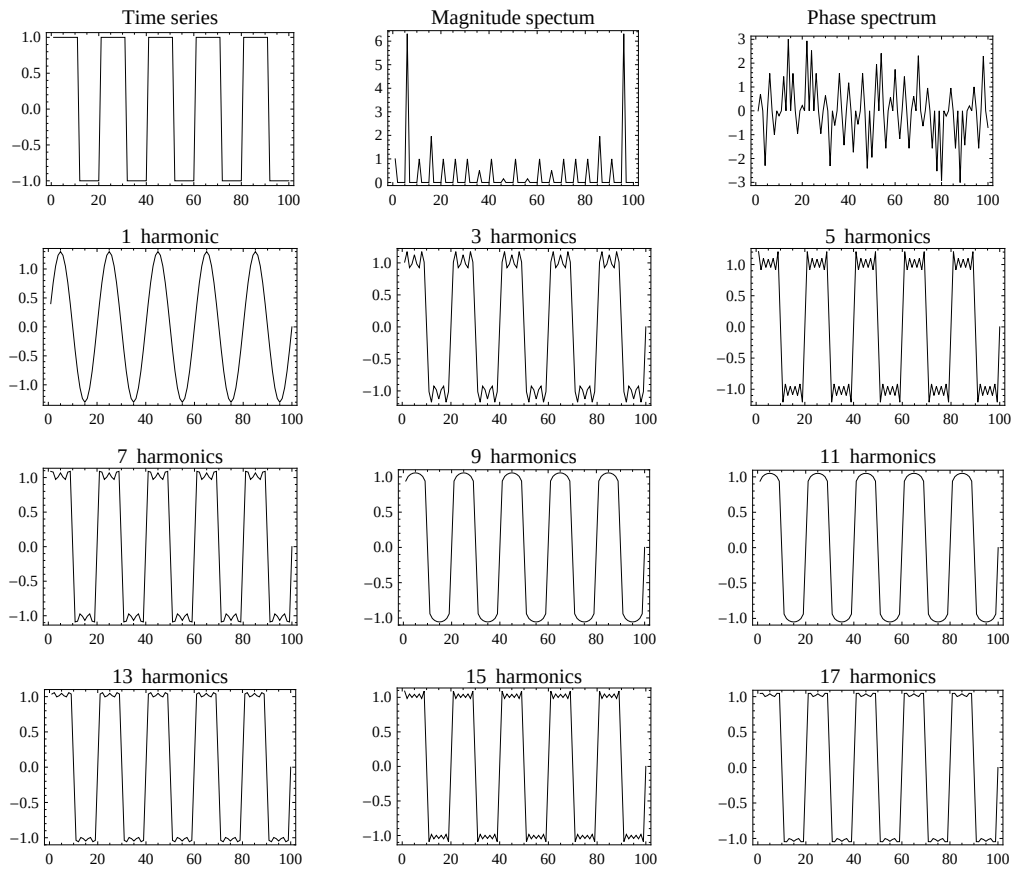
Έτσι, είναι δυνατόν να αναπαραχθούν τα αρχικά δεδομένα από την έξοδο αυτή. Μπορεί κανείς να δει στο Σχήμα 1.6, ακόμα και με χρήση ενός υποσυνόλου των αρμονικών¹ είναι δυνατόν να προσεγγιστεί ένα τετραγωνικό σχήμα (περίπτωση συνάρτησης που θεωρείται ‘δύσκολη’ να αναλυθεί σε ημιτονοειδή σήματα).

Θεώρημα δειγματοληψίας Nyquist–Shannon

Οι Harry Theodor Nyquist και Claude Elwood Shannon ανακάλυψαν – ανεξάρτητα ο ένας από τον άλλο – το ‘θεώρημα δειγματοληψίας’ σύμφωνα με το οποίο,

“Ο ρυθμός δειγματοληψίας που απαιτείται για να αναπαραχθεί ακριβώς ένα σήμα από τα δείγματά του, ο **ρυθμός Nyquist**, είναι μεγαλύτε-

¹ Με αυτόν τον τρόπο είναι δυνατός ο περιορισμός του όγκου πληροφορίας όταν ένα επιθυμητό επίπεδο προσέγγισης έχει επιτευχθεί — καθιστώντας το μετασχηματισμό Fourier ισχυρό εργαλείο απωλεστικής συμπίεσης



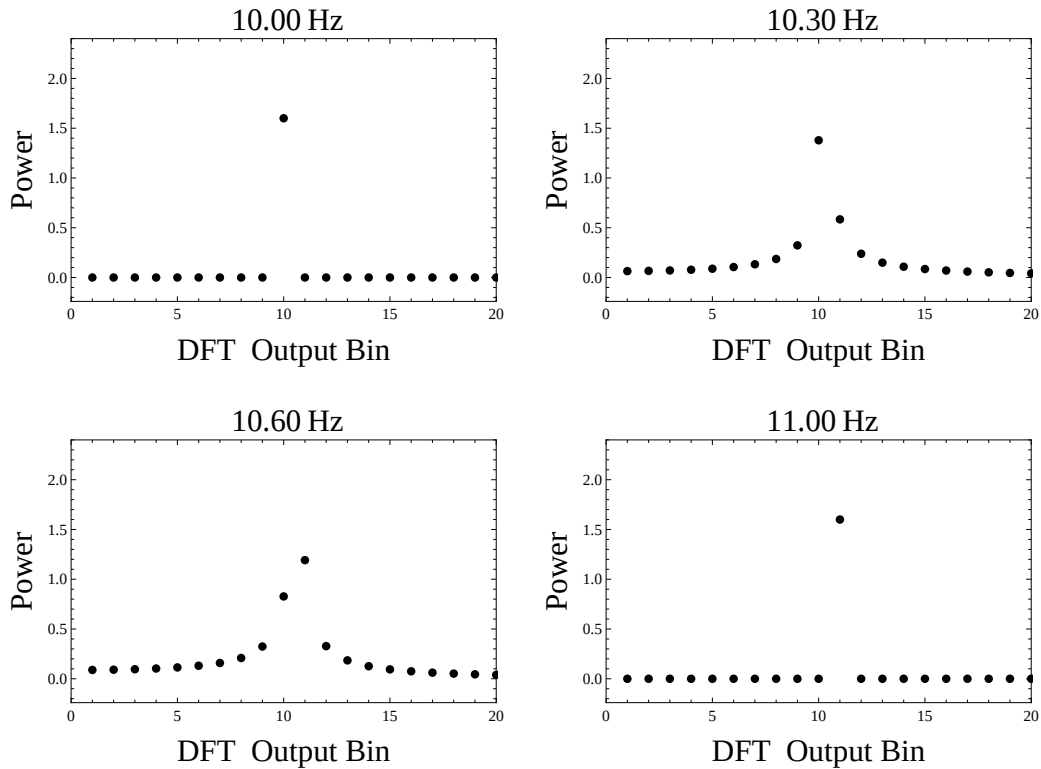
Σχήμα 1.6: Ένα παράδειγμα ανάλυσης ενός τετραγωνικού σήματος (πρώτη γραμμή, πρώτη στήλη) στο φάσμα του. Το πλάτος και η φάση κάθε αρμονικής δίνεται στην πρώτη γραμμή. Καθώς προσθέτουμε όλο και περισσότερες αρμονικές, έχουμε καλύτερη προσέγγιση στο αρχικό σήμα (σύνθεση)

ρος από δύο φορές τη μέγιστη συχνότητα για την οποία ο μετασχηματισμός Fourier του σήματος είναι μη μηδενικός, η οποία αποκαλείται **συχνότητα Nyquist** ή συχνότητα αναδίπλωσης.”

Όταν το κριτήριο Nyquist ικανοποιείται, το φαινόμενο του aliasing (μετάπτωση στο φάσμα) εμφανίζεται. Παράδειγμα του είναι η εμφάνιση ενός τροχού να περιστρέφεται ανάποδα από την πραγματική διεύθυνση όταν η ταχύτητα φτάσει σε μία ορισμένη τιμή. Η δειγματοληψία της κάμερας ή των ματιών μας δεν είναι αρκετά συχνή ώστε να αντιληφθούμε την πραγματική περίοδο.

Εξαιτίας του φαινομένου αυτού,

- η αναζήτηση για περιοδικότητες σε μία σειρά N αριθμών περιορίζεται από



Σχήμα 1.7: Παράδειγμα της φασματικής διάχυσης στον διακριτό μετασχηματισμό Fourier. Τα τέσσερα σχήματα αντιστοιχούν στους μετασχηματισμούς ενός σήματος ημιτονοειδούς μορφής με συχνότητα που αναγράφεται στα σχήματα. Οι ακέραιες συχνότητες δεν παρουσιάζουν φασματική διάχυση και συντελούν στην εμφάνιση ενός bin μέγιστης ισχύος

2 ως $N/2$ φορές την περίοδο δειγματοληψίας

- οι μισοί από τους συντελεστές Fourier μίας χρονοσειράς πραγματικών αριθμών μπορούν να απορριφθούν ($N/2$ ως $N - 1$)

Φασματική διάχυση

Κάθε στοιχείο του DFT, ή ‘bin συχνότητας’, ενός σήματος αντιστοιχεί σε ένα πολλαπλάσιο μίας βασικής συχνότητας. Αν το προς ανάλυση σήμα περιλαμβάνει μία ημιτονοειδή συνιστώσα συχνότητας ίση με κάποιο bin συχνότητας, τότε θα δώσει μόνο μία μη μηδενική τιμή στον DFT: καμία άλλη συχνότητα δε θα ανιχνευθεί. Τι συμβαίνει όμως στην περίπτωση που η συχνότητα δεν είναι πολλαπλάσια

της βασικής; Όλα τα bin συχνότητας λαμβάνουν μη μηδενική τιμή και μία κωδωνοειδής κατανομή σχηματίζεται στην περιοχή κοντά στη συχνότητα του σήματος. Κατά συνέπεια, το μέγιστο εντοπίζεται κοντά στην πραγματική συχνότητα και η ισχύς διαχέεται με αποτέλεσμα τη μείωσή του κατά απόλυτη τιμή (Σχήμα 1.7). Το φαινόμενο αυτό αποκαλείται **φασματική διάχυση**.

Αναπαράσταση DFT σε πίνακα

Η Εξίσωση (1.26) μπορεί να γραφτεί σε μορφή πινάκων ως

$$\begin{pmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{pmatrix} = \begin{pmatrix} w^0 & w^0 & w^0 & \dots & w^0 \\ w^0 & w^1 & w^2 & \dots & w^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w^0 & w^{N-1} & w^{2(N-1)} & \dots & w^{(N-1)^2} \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{pmatrix} \quad (1.28)$$

όπου

$$w = e^{-i\frac{2\pi}{N}} \quad (1.29)$$

Εκμεταλλευόμενοι τις συμμετρίες στον μέσο πίνακα, οδηγούμαστε σε βελτιστοποιημένους αλγορίθμους για τον υπολογισμό των DFT, όπως ο ‘Radix–2 Fast Fourier Transform’ που θα αναλυθεί εδώ.

1.4.4 Radix–2 Fast Fourier Transform

“Διαίρει και βασίλευε”

Η διαίρεση ενός προβλήματος σε υποπροβλήματα είναι συχνά πολύ αποδοτική. Ας πάρουμε το απλό παράδειγμα μίας απλής αναζήτησης ενός ονόματος σε έναν τηλεφωνικό κατάλογο. Διαβάζοντας τις εγγραφές μία–προς–μία μπορεί να δώσει αποτελέσματα μετά από ώρες ή μέρες. Αντιθέτως όλοι μας, ενστικτωδώς κάνουμε μία έξυπνη επιλογή των σελίδων στις οποίες αναζητούμε. Ανοίγουμε το βιβλίο σε μία σελίδα και ελέγχουμε αν το όνομα είναι πριν ή μετά από το τρέχον σημείο. Έπειτα συνεχίζουμε την έρευνα στο κατάλληλο μισό του βιβλίου και ούτω καθ’ εξής. Η αναζήτηση θα έχει τελειώσει μέσα σε λίγα δευτερόλεπτα.

Η ίδια λογική μπορεί να εφαρμοστεί στον υπολογισμό του διακριτού μετασχηματισμού Fourier. Όταν το N είναι άρτιο, τότε η Εξίσωση (1.26) μπορεί να χωριστεί σε δύο μέρη (περιττά και άρτια n):

$$X_p = \sum_{n=0}^{\frac{N}{2}-1} x_{2n} e^{-i \frac{2\pi}{N} (2n)p} + \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1} e^{-i \frac{2\pi}{N} (2n+1)p} = \quad (1.30)$$

$$= \sum_{n=0}^{\frac{N}{2}-1} x_{2n} e^{-i \frac{2\pi}{N/2} np} + e^{-i \frac{2\pi}{N} p} \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1} e^{-i \frac{2\pi}{N/2} np} = \quad (1.31)$$

$$= A_p + w^p B_p \quad (1.32)$$

όπου η έκφραση

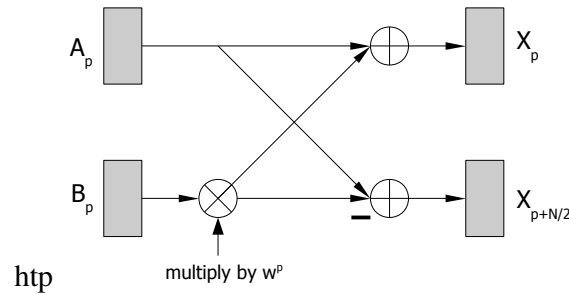
$$A_p = \sum_{n=0}^{\frac{N}{2}-1} x_{2n} e^{-i \frac{2\pi}{N/2} np} \quad (1.33)$$

$$B_p = \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1} e^{-i \frac{2\pi}{N/2} np} \quad (1.34)$$

είναι προφανώς ο DFTs των μεγέθους $\frac{N}{2}$ ακολουθιών

- $(x_0, x_2, \dots, x_{N-2})$
- $(x_1, x_3, \dots, x_{N-1})$

Η διαίρεση του προβλήματος σε δύο μικρότερα μέρη βελτιστοποιεί τον αλγόριθμο εξ' αιτίας του προηγουμένως αναφερόμενου υπολογιστικού κόστους ($O(N^2)$) του DFT. Υπολογίζοντας δύο μισού μεγέθους DFT μειώνουμε το πλήθος των πράξεων κατά έναν παράγοντα 2: $\left(\frac{N}{2}\right)^2 + \left(\frac{N}{2}\right)^2 = \frac{N^2}{2}$.



Σχήμα 1.8: Το ‘διάγραμμα πεταλούδας του FFT’. Οι ίσου μεγέθους πίνακες A_p and B_p συνδυάζονται για να δώσουν περιττής και άρτιας θέσεως στοιχεία του DFT, X_p

Συμμετρία

Εκμεταλλευόμενοι το γεγονός ότι ο DFT είναι περιοδικός με περίοδο $\frac{N}{2}$, οδηγούμαστε σε μία ακόμη απλοποίηση:

$$X_p = A_p + w^p B_p \quad (1.35)$$

$$X_{p+\frac{N}{2}} = A_p - w^p B_p \quad (1.36)$$

Το διάγραμμα ροής αυτής της διαδικασίας (Σχήμα 1.8) λέγεται ‘διάγραμμα πεταλούδας του FFT’ και αντανακλά τον συνδυασμό αποτελεσμάτων δύο DFT πάνω σε δύο μέρη μίας χρονοσειράς, σε έναν DFT του συνόλου των δεδομένων.

Radix-2: στρατηγική και διάγραμμα ροής

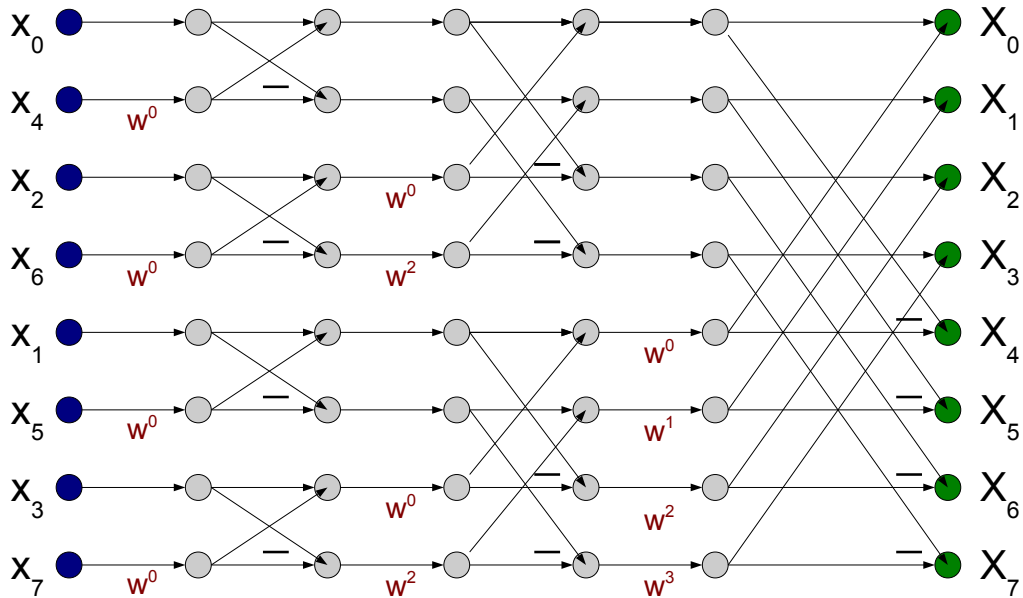
Αν $N = 2^M$, τότε η στρατηγική του ‘διαίρει και βασίλευε’ μπορεί να εφαρμοστεί M φορές, ανάγοντας τον υπολογισμό του DFT μιας μεγάλης χρονοσειράς στον υπολογισμό πολλών DFT 2 σημείων. Καθώς τα δεδομένα χωρίζονται σε περιττές και άρτιες τιμές, η είσοδος πρέπει να ανακατευτεί καταλλήλως. Αυτό γίνεται με την αναστροφή των bit¹ των αριθμών των θέσεων (Πίνακας 1.2).

Το συνολικό διάγραμμα ροής του Radix-2 FFT στην περίπτωση 8 στοιχείων δίνεται στο Σχήμα 1.9.

¹ Στις διάφορες πραγματοποιήσεις του FFT, η αναστροφή των bit μπορεί να γίνει είτε απευθείας είτε εμμέσως

Πίνακας 1.2: Παράδειγμα αναστροφής bit των δεικτών σε έναν πίνακα 8 στοιχείων

Θέση (decimal)	Θέση (δυαδικό)	Αναστροφή bit	Νέα θέση
0	000	000	0
1	001	100	4
2	010	010	2
3	011	110	6
4	100	001	1
5	101	101	5
6	110	011	3
7	111	111	7



Σχήμα 1.9: Διάγραμμα ροής του FFT. Η είσοδος (μπλε) ανακατεύεται (αναστροφή bit) και μία σειρά διατάξεων ‘πεταλούδας FFT’ τη μετατρέπει στον διακριτό μετασχηματισμό Fourier (πράσινο). Παρατηρείστε ότι το διάγραμμα αυτό διαφέρει από εκείνο στο Σχήμα 1.4: **i.** στοιχεία και όχι ομάδες στοιχείων υπόκεινται στους υπολογισμούς, **ii.** δεν υπάρχουν μεταθέσεις στοιχείων αλλά πολλαπλασιασμοί με δυνάμεις του w (κόκκινο) ή με -1 (μαύρα σύμβολα μείον)

Ιστορικό σημείωμα

Δύο δεκαετίες μετά τη δημοσίευση του πρώτου FFT (Cooley & Tukey, 1965) ανακαλύφθηκε πώς ο διάσημος Γερμανός μαθηματικός, Carl F. Gauss, είχε εφεύ-

ρει τη μέθοδο FFT το 1805, νωρίτερα δηλαδή από τη μελέτη του Γάλλου μαθηματικού, Joseph B. J. Fourier για τις σειρές Fourie (Heideman et al., 1985).

1.5 Ιδιότητες συγκεκριμένων τύπων καμπυλών φωτός

Σε αυτήν την ενότητα θα κάνουμε επισκόπηση τριών τύπων περιοδικών μεταβλητών αστερών και των καμπυλών φωτός τους. Η επιλογή αυτών των ομάδων δεν έγινε τυχαία, καθώς ταιριάζουν σε δύο ακραίες περιπτώσεις και σε μία ενδιάμεση. Η μία κατηγορία είναι οι **Κηφείδες μεταβλητοί αστέρες** από χαρακτηρίζονται από σχεδόν ημιτονοειδή καμπύλη φωτός. Η άλλη, είναι οι **αστέρες με διαβαίνοντες εξωπλανήτες**, όπου στόχος είναι ο εντοπισμός μικρών ορθογώνιων παλμών (κατά προσέγγιση). Ως ‘γέφυρα’ που ενώνει τα χαρακτηριστικά των δύο αυτών ομάδων, θεωρούμε κατάλληλη την επιλογή των **διπλών εκλειπτικών συστημάτων** που παρουσιάζουν ημιτονοειδείς ή τετραγωνικές μορφές μικρού ή μεγάλου κύκλου εργασίας — αναλόγως τύπου και ιδιοτήτων.

1.5.1 Κηφείδες

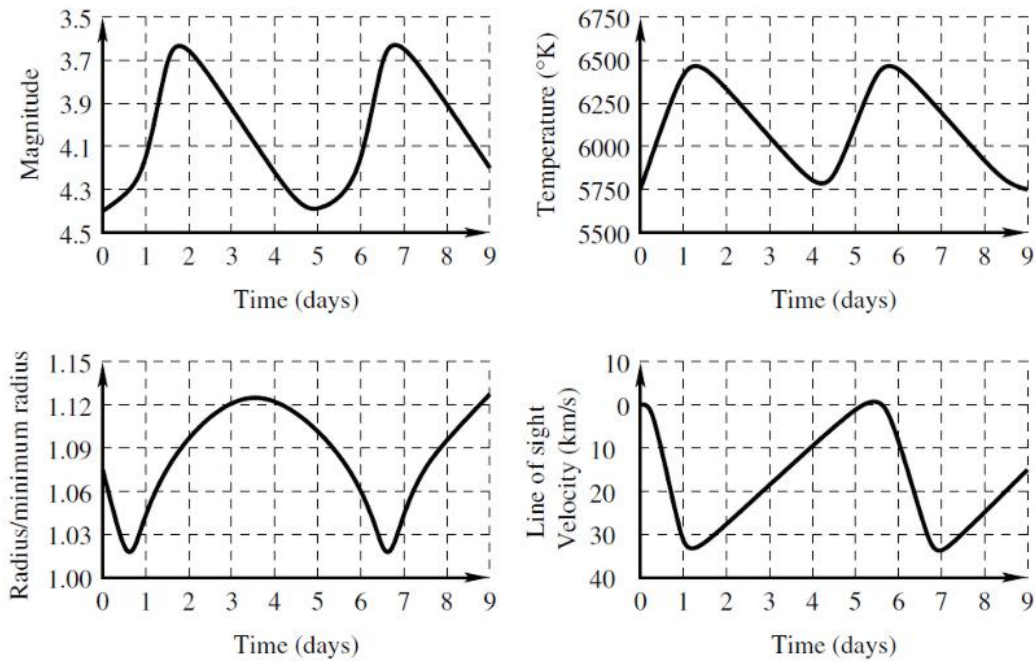
Γενικές πληροφορίες

Η ομάδα των μεταβλητών αυτών ονομάστηκε από τον πρωτότυπο – *δ Cephei* – και είναι μεγάλης ιστορικής σημασίας για την αστρονομία και αστροφυσική. Ο κύκλος αναπάλσεων ενός Κηφείδη προδίδει το απόλυτο μέγεθός του (σχέση περιόδου – λαμπρότητας των Κηφείδων) το οποίο συνδυαζόμενο με μία μέτρηση του φαινομένου μεγέθους οδηγεί σε μία εκτίμηση της απόστασης το αστέρα ή του γαλαξία στον οποίο ανήκει (Shu, 1982).

Οι Κηφείδες είναι ‘ενδογενείς μεταβλητοί’ καθώς η μεταβλητότητά τους προκαλείται από εσωτερικές φυσικές διεργασίες. Ο Sir Arthur Eddington υπέδειξε έναν μηχανισμό (‘βαλβίδα του Eddington’) για την ανάπαυση των αστερών λόγω

μηχανικής αστάθεια που αργότερα χρησιμοποιήθηκε για να εξηγήσει τη συμπεριφορά των Κηφείδων. Με απλά λόγια, αυτοί οι αστέρες είναι σα μπαλόνια που συμπιέζονται και εκτονώνονται (radial pulsation mode), αλλάζοντας τη λαμπρότητά του καθώς η ακτινοβολία αλληλεπιδρά με το υλικό τους (Percy, 2007).

Ιδιότητες καμπυλών φωτός



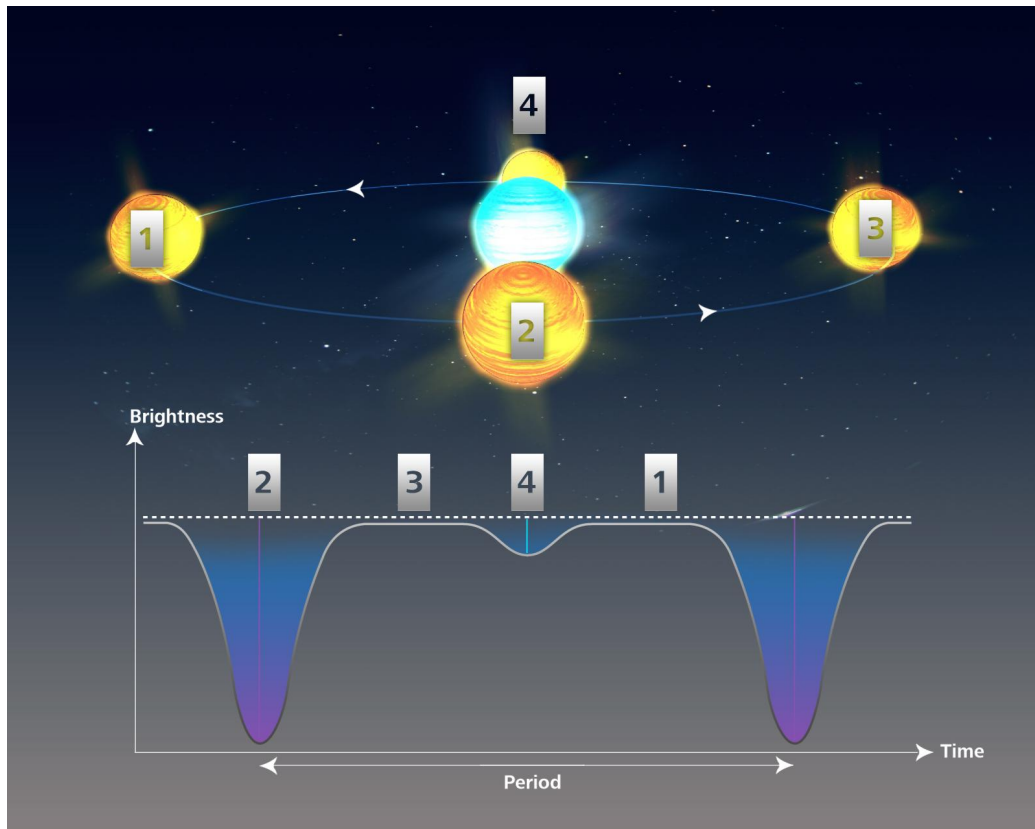
Σχήμα 1.10: Το μέγεθος (πάνω αριστερά), η θερμοκρασία (πάνω δεξιά), η ακτίνα (κάτω αριστερά) και η ακτινική ταχύτητα (κάτω δεξιά) του δ του Κηφέα (Percy, 2007).

Το προφίλ ενός Κηφείδη μεταβλητού μοιάζει με λοξές ημιτονοειδείς κορυφές (Σχήμα 1.10). Τα μέγιστα και τα ελάχιστα δε συμβαίνουν όταν η ακτίνα βρίσκεται στις ακραίες τιμές καθώς μεταβάλλεται ταυτόχρονα και η θερμοκρασία. Θυμηθείτε την γνωστή σχέση

$$L = 4\pi R^2 \sigma T^4 \quad (1.37)$$

όπου L είναι η λαμπρότητα, R η ακτίνα και T η ενεργός θερμοκρασία του αστέρα.

1.5.2 Διπλοί εκλειπτικοί αστέρες



Σχήμα 1.11: Αναπαράσταση του τρόπου παραγωγής των καμπυλών φωτός των διπλών εκλειπτικών. Credit: German Aerospace Center (*Deutsches Zentrum für Luft- und Raumfahrt e.V.*)

Γενικές πληροφορίες

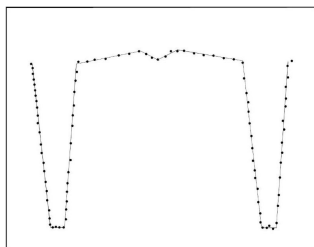
Δύο αστέρες που περιφέρονται γύρω από ένα κοινό κέντρο μάζας σχηματίζουν ένα διπλό αστέρα — ή διπλό σύστημα αστερών πιο σωστά. Ένας **διπλός εκλειπτικός αστέρας** έχει το επίπεδο τροχιάς των μελών του πολύ κοντά στην ευθεία παρατήρησης του από τη Γη. Έτσι, συμβαίνουν αμοιβαίες εκλείψεις (όπως φαίνεται στον παρατηρητή), που εντοπίζονται ως περιοδικές πτώσης του μεγέθους του αστέρα (§ 1.5.2) (Warner, 2006).

Είναι ‘εξωγενής μεταβλητοί’ καθώς η μεταβλητότητά τους δεν οφείλεται σε εσωτερικές μεταβολές στους αστέρες. Φυσικά ένα ή και τα δύο μέλη μπορούν να είναι και ενδογενώς μεταβλητοί.

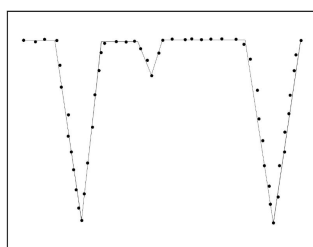
Τύποι διπλών εκλειπτικών

Οι βασικοί τύποι τέτοιων συστημάτων είναι Algol, Beta Lyræ(β Lyr) και W UrsæMajoris (W UMa), και αντιστοιχούν σε διαφορετική μηχανική εικόνα. Στους Algol διπλούς, τα μέλη είναι καλά διαχωρισμένα, ενώ στους Beta Lyræ και W UMa έχουμε συστήματα των οποίων οι αστέρες είναι πολύ κοντά μεταξύ τους. Οι καμπύλες φωτός των ‘κλειστών διπλών’ παρουσιάζουν μία έντονη παραμόρφωση (“ωσειδής μορφή”) λόγω των παλιρροϊκών δυνάμεων μεταξύ των μελών (Σχήμα 1.12(γ')). Στην περίπτωση των W UMa, ένα ή και τα δύο αστέρια έχουν επεκταθεί πέρα από το ‘όριο Roche’ και η βίαιη φύση τους μπορεί να προκαλέσει την αλλαγή του προφίλ της περιοδικότητας δραματικά ακόμα και μέσα σε λίγους κύκλους.

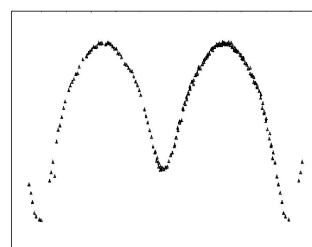
Ιδιότητες καμπυλών φωτός διπλών εκλειπτικών



(α') Φαινόμενα ‘αντανάκλασης’ σε έναν διπλό τύπου Algol



(β') Η επίδραση της εκκεντρότητας



(γ') Κλειστό σύστημα εκλειπτικών διπλών

Σχήμα 1.12: Διάφορα στοιχεία και επιδράσεις στις καμπύλες φωτός των διπλών εκλειπτικών. Εικόνες από Warner (2006).

Η μορφή των καμπυλών φωτός εξαρτάται από πολλούς παράγοντες που έχουν διαφορετικές επιδράσεις στην εικόνα τους:

- **Σχετικό μέγεθος:** Αν οι αστέρες έχουν περίπου το ίδιο απόλυτο μέγεθος (ή φαινόμενο λόγω της ίδιας απόστασης από εμάς), τότε η πτώση της έντασης

του φωτός σε κάθε έκλειψη θα είναι ίδια. Στις περισσότερες περιπτώσεις υπάρχει παρατηρούμενη διαφορά στη λαμπρότητα των μελών του συστήματος — οδηγώντας σε μία αλληλουχία μεγάλων και μικρών πτώσεων στην καμπύλη φωτός όταν ο λαμπρότερος ή ο αμυδρότερος αστέρας εκλείπεται αντίστοιχα. Έτσι έχουμε **πρωτεύοντα** και **δευτερεύοντα** ελάχιστα.

- **Ενδογενής μεταβλητότητα:** Ένα ή και τα δύο από τα μέλη είναι δυνατόν να είναι μεταβλητοί αστέρες άλλου τύπου.
- **Εκκεντρότητα:** Σε μία κυκλική τροχιά οι εκλείψεις συμβαίνουν με ίση μεταξύ τους απόσταση στο χρόνο. Από την άλλη, η εκκεντρότητα μπορεί να μεταβάλει τη διάρκεια ανάμεσα στα ελάχιστα. Φυσικά, το φαινόμενο εξαρτάται και από τη διεύθυνση της τροχιάς σε σχέση με τον παρατηρητή. Για παράδειγμα, αν η ευθεία παρατήρησης είναι περίπου παράλληλη στο μεγάλο ημιάξονα της τροχιάς, τότε δε θα παρουσιαστεί μετακίνηση των ελαχίστων (Σχήμα 1.12(β')).
- **Αμαύρωση χείλους:** ο δίσκος ενός αστέρα δεν είναι ομοιόμορφα φωτεινός. Κοιτώντας στο κέντρο του, βλέπουμε σε βαθύτερα και θερμότερα στρώματα της αστρικής ατμόσφαιρας απ' ότι όταν κοιτούμε στις άκρες του δίσκου. Έτσι, το οπτικό βάθος εξαρτάται από την απόσταση του εκάστοτε σημείου από το κέντρο, κάτι που συνεπάγεται (i) ομαλότερη μείωση της λαμπρότητας του συστήματος και ανεπίδεδα ελάχιστα κατά τις εκλείψεις. Υπάρχουν διάφορα μοντέλα περιγραφής του φαινομένου και συνήθως λαμβάνονται υπόψη δύο συντελεστές αμαύρωσης χείλους.
- **Κλίση:** όταν η ευθεία παρατήρησης ενός διπλού συστήματος είναι αρκετά παράλληλη με το επίπεδο της τροχιάς του, το μικρότερο σε ακτίνα μέλος θα εκλείπεται πλήρως παρουσιάζοντας ένα επίπεδο κατώφλι στο δευτερεύον ελάχιστο. Σε αντίθετη περίπτωση, κανένας αστέρας δε μπορεί να επισκιαστεί πλήρως, με αποτέλεσμα καμπύλα ελάχιστα αλλά και μικρότερα (κάποιες φορές το δευτερεύον ελάχιστο δεν είναι παρατηρήσιμο).
- **Φαινόμενο 'ανάκλασης':** στο ενδιάμεσο ανάμεσα στα ελάχιστα, κάποιες φορές παρατηρείται μία ελαφρά αύξηση και έπειτα μείωση του μεγέθους

του συστήματος με την κορυφή να βρίσκεται κοντά στο δεύτερο ελάχιστο. Αιτία είναι η επιπλέον θέρμανση της μίας πλευράς του μικρότερου μέλους από το μεγαλύτερο, που το καθιστά φωτεινότερο (Σχήμα 1.12(α')).

- **Αμαύρωση βαρύτητας:** για κλειστά συστήματα διπλών, οι παλιρροϊκές δυνάμεις μεταβάλλουν το σχήμα τους σε ελλειψοειδές των οποίων τα άκρα είναι ψυχρότερα, οπότε και λιγότερο φωτεινά. Έτσι, η αλλαγή της διεύθυνσής τους καθώς περιφέρονται σχεδόν κλειδωμένα το ένα απέναντι στο άλλο, παρουσιάζεται σαν μία επιπλέον μεταβλητότητα.

1.5.3 Διαβάσεις εξω-ηλιακών πλανητών

Γενικές πληροφορίες

Από τα τέλη της δεκαετίας του '80 και τις πρώτες ενδείξεις για ύπαρξη πλανητών σε άλλους αστέρες¹, και την πρώτη ξεκάθαρη ανίχνευση ενός εξωπλανήτη περιφερόμενου του αστέρα 51 Peg (Mayor & Queloz, 1995), 885 επιβεβαιωμένοι και χιλιάδες υποψήφιοι πλανήτες έχουν ανακαλυφθεί².

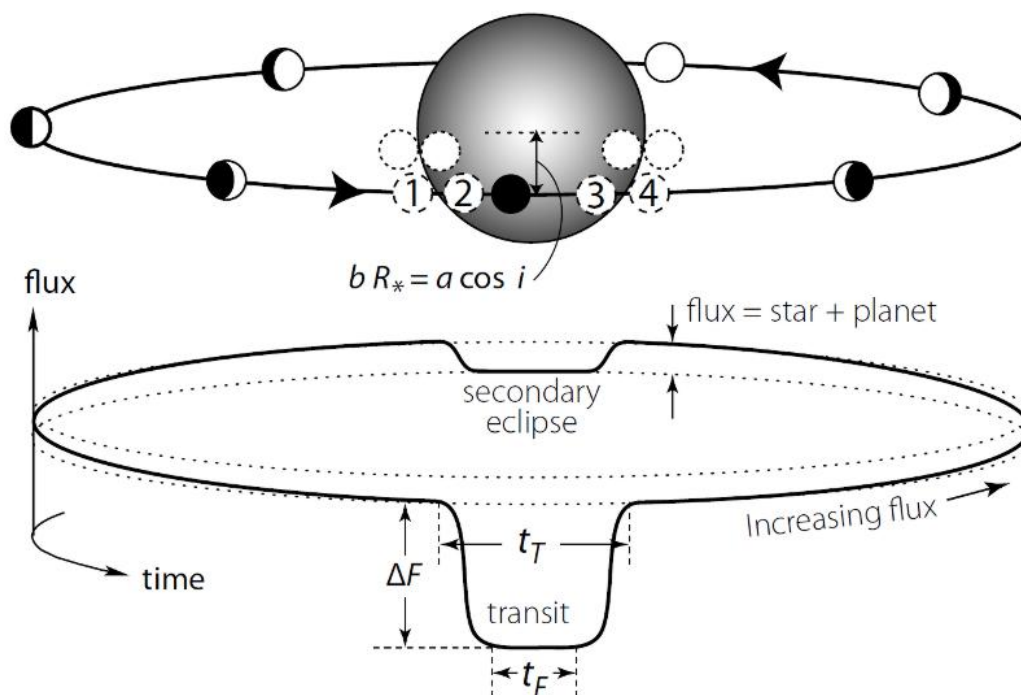
Υπάρχουν πολυάριθμες επιτυχείς ή προτεινόμενες τεχνικές για την αναζήτηση εξωπλανητών. Η μέθοδος των **διαβάσεων** είναι η πιο πετυχημένη μετά την **μέθοδο των ακτινικών ταχυτήτων** και εκμεταλλεύεται το σχήμα της καμπύλης φωτός του αστέρα γύρω από τον οποίο περιφέρεται ο υποτιθέμενος εξωπλανήτης. Η βασική ιδέα είναι ότι μπορούμε να μετρήσουμε μικρές μειώσεις του φωτός του αστέρα σε περίπτωση που ένας πλανήτης περνάει μπροστά από τον πρώτο όπως τον παρατηρούμε από τη Γη (Perryman, 2011).

Ιδιότητες της καμπύλης φωτός

Η καμπύλη φωτός του αστέρα που περιφέρεται από έναν διαβαίνοντα εξωπλανήτη μπορεί να θεωρηθεί ως μία παραλλαγή των διπλών εκλειπτικών συστη-

¹ Φιλόσοφοι και λόγιοι της αρχαιότητας και της αναγέννησης υπέθεταν ότι τα αστέρα ήταν μακρινοί Ήλιοι που θα μπορούσαν να έχουν τις δικές τους Γαίες

² Στοιχεία της 7ης Μαΐου του 2013. Πηγή: The Extrasolar Planets Encyclopaedia (www.exoplanet.eu)

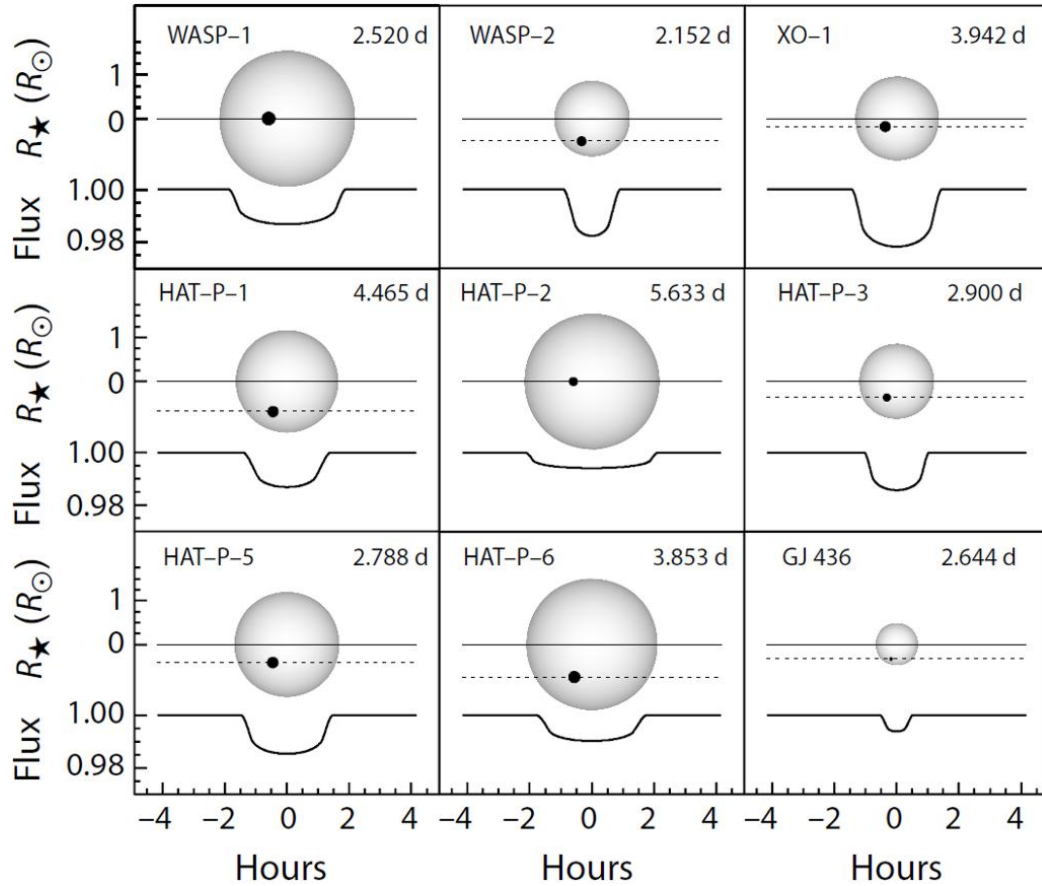


Σχήμα 1.13: Τριδιάστατη αναπαράσταση μία διάβασης εξωπλανήτη και η επίδραση των διαφόρων φάσεων στη μορφή της καμπύλης φωτός του συνοδού αστέρα (Perryman, 2011).

μάτων. Παρά το ότι ο συνοδός είναι πλανήτης και όχι αστέρα, οι διαφορές είναι περισσότερο ποσοτικές και όχι ποιοτικές (Σχήμα 1.13). Σε αντίθεση με την Σχήμα 1.11,

- το μικρό μέγεθος των πλανητών δημιουργεί μικρότερα ελάχιστα στην ένταση κατά την πρωτεύουσα έκλειψη και τείνει να δημιουργεί πιο επίπεδα προφίλ κατά το ελάχιστο, σε σχέση με τους διπλούς εκλειπτικούς
- η ακτινοβολία του πλανήτη είναι εκατομμύρια ή δισεκατομμύρια φορές λιγότερο ισχυρή από την αστρική στην οπτική περιοχή του H/M φάσματος, καθιστώντας τη δευτερεύουσα έκλειψη μη ανιχνεύσιμη. Έτσι, η εκκεντρότητα μελετάται από προσεκτική παρατήρηση του προφίλ της διάβασης και όχι από τη χρονομέτρηση ανάμεσα στα ελάχιστα
- το φαινόμενο της αμαύρωσης του χείλους είναι παρόν και εδώ (Σχήμα 1.14), κάνοντας τα ελάχιστα καμπύλα. Η κλίση της τροχιάς σε σχέση με την ευθεία παρατήρησης παίζει σημαντικό ρόλο εδώ

- το φαινόμενο της ανάκλασης είναι ασθενές αλλά έχει παρατηρηθεί σε κάποιες περιπτώσεις ¹



Σχήμα 1.14: Διάφορα προφίλ διαβάσεων. Τα μεγέθη πλανητών — αστέρων δίνονται σε κλίμακα. Μπορούμε να διακρίνουμε την επίδραση του μεγέθους των πλανητών, της κλίσης της τροχιάς και της αμαύρωσης του χείλους στην καμπύλη φωτός (Perryman, 2011).

1.6 Στόχοι της εργασίας

Για να ελέγξουμε και να συγκρίνουμε τη λειτουργία των αλγορίθμων FFT και FFA σε αστρονομικά δεδομένα θα πραγματοποιήσουμε μία σειρά από τεστ και

¹ από τις αποστολές CoRoT και Kepler, ενώ κατάφεραν και να εντοπίσουν εξωπλανήτες εκμεταλλευόμενοι το φαινόμενο αυτό

θα επιχειρήσουμε να απαντήσουμε σε κάποια ερωτήματα:

1. Πραγματοποίηση των αλγορίθμων και έλεγχος καλής λειτουργίας τους
2. Είναι τόσο χρονοβόροι όσο περιμένουμε;
3. Ποιά είναι η επιτυχία τους στην εύρεση περιοδικοτήτων σε χρονοσειρές με διάφορες παραμέτρους: σχήμα, περίοδος, επίπεδα θορύβου;
4. Μπορούν να χρησιμοποιηθούν στην ανίχνευση περιόδων των τριών αντιπροσωπευτικών τάξεων αντικειμένων (Κηφειδες, διπλοί εκλειπτικοί και διαβάσεις εξωπλανητών);

“— Γουότσον, σημειώνεις πρόοδο. Τα πηγές πράγματι περίφημα. Είναι αλήθεια ότι σου ξέφυγε κάθε τι σημασίας, αλλά πέτυχες διάνα στη μέθοδο...”

Σέρλοκ Χολμς στο ‘Υπόθεση Ταυτότητας’ του Σερ Άρθουρ Κόναν Ντόιλ

2

Μεθοδολογία

2.1 Προγραμματισμός

2.1.1 Συμβάσεις

Ανάμεσα στις διάφορες γλώσσες προγραμματισμού που χρησιμοποιούνται συχνά στην αστροφυσική (Fortran, C/C++, IDL, Perl, κτλ.), επιλέξαμε την ANSI C (or C89/C90) για συμβατότητα, εύκολη ανάγνωση και λόγω εξοικείωσης του γράφοντα με αυτήν. Σε αντίθεση με την αντικειμενοστρεφή C++, οι αλγόριθμοι που αναπτύχθηκαν εδώ μπορούν εύκολα να μεταγλωττιστούν σε άλλες γλώσσες δομημένου προγραμματισμού.

Η τρέχουσα έκδοση του λογισμικού που δημιουργήσαμε, χειρίζεται τα αστρονομικά δεδομένα ως πίνακες μεταβλητών κινητής υποδιαστολής (μέγεθος μίας υπολογιστικής λέξης, που σημαίνει 32-bit στους περισσότερους σύγχρονους υπολογιστές). Από την άλλη, οι χρόνοι αποθηκεύονται με διπλοί ακρίβεια (δύο λέξεις,

64-bit) καθώς αποφασίσαμε να χρησιμοποιήσουμε μέρες BJD ως μονάδες: πολλές χρονοσειρές λαμβάνονται με ακρίβεια και αποστάσεις των τάξεων των ms (π.χ. για pulsar).

Ένα πλεονέκτημα των κινητής υποδιαστολής δεδομένων είναι ότι δε χρειάζεται να ελέγχουμε για υπέρβαση της μέγιστης/ελάχιστης τιμής κατά την άθροιση, κάτι που θα οδηγούσε σε αλλαγή προσήμου¹. Επίσης, η είσοδος και έξοδος ήταν σε μορφή ASCII.

2.1.2 Πραγματοποίηση κώδικα FFT, FFA και αναδίπλωσης

Εδώ βρίσκεται μία περιγραφή των βασικών αλγορίθμων του λογισμικού. Η είσοδο των FFT και FFA είναι πίνακες παρατηρήσεων (bins). Οι συναρτήσεις αναδίπλωσης τροφοδοτούνται και με τους χρόνους παρατήρησης.

FFT

Ο Fast Fourier Transform εκτελείται με τη συνεργασία δύο συναρτήσεων. Η συνάρτηση FFT:

- Δεσμεύει χώρο στη μνήμη για το buffering
- Καλεί τη συνάρτηση RECURSIVE_FFT που με τη σειρά της
 - καλεί τον εαυτό της δύο φορές— για κάθε μισό της εισόδου της, μέχρι να φτάσει σε πίνακες δύο στοιχείων
 - συνδυάζει τα δύο μισά, με τον τρόπο της ‘πεταλούδας FFT’, και επιστρέφει το αποτέλεσμα

Η διαδικασία τελικά καταλήγει στο διακριτό μετασχηματισμό της εισόδου

- Η έξοδος αποθηκεύεται σε ένα αρχείο αναφοράς
- ... και ελέγχεται για το μέγιστο το οποίο θεωρούμαι και πιο πιθανή τιμή για την περίοδο του σήματος

¹ Για παράδειγμα μία μεταβλητή τύπου char στη γλώσσα προγραμματισμού C θα περάσει από την τιμή 127 στην τιμή -128 με πρόσθεση του 1

- Καθώς η φασματική διάχυση δεν αναλύεται εδώ, μία αβεβαιότητα υπολογίζεται:
 - i_{max} είναι η θέση του μεγίστου του DFT
 - η ελάχιστη και μέγιστη περίοδος που θα έδιναν μέγιστο σε αυτήν τη θέση είναι:

$$\frac{\text{διάρκεια εισόδου}}{i_{max} \pm 0.5}$$

- ο μέσος τους λαμβάνεται ως η περίοδος του σήματος
- το μισό της διαφοράς τους αποθηκεύεται ως σφάλμα διακριτοποίησης

FFA

Ο Fast Folding Algorithm πραγματοποιείται από τη συνάρτηση FFA που κάνει χρήση της δομής FFA_RESULT για να επιστρέφει τα αποτελέσματα. Η τελευταία, περιλαμβάνει την ανιχνευμένη περίοδο `period`, την ανάλυση ή απόσταση μεταξύ δύο περιόδων στην έξοδο `dp`, το μέτρο σημαντικότητας `significance` και τη σημαία `outcome` για επιτυχία ή αποτυχία.

Ο αλγόριθμος,

- υπολογίζει τον αριθμό των βασικών περιόδων που μπορούν να χωρέσουν σε όλη την έκταση της χρονοσειράς και τον μειώνει στη μέγιστη δύναμη του 2. Αυτός είναι και ο αριθμός των υποπινάκων που θα καταλήξουν ως έξοδος για διαφορετικές περιόδους (αναδιπλώσεις)
- δεσμεύει μνήμη για το buffering
- εφαρμόζει τον FFA κατά βήματα δουλεύοντας στο ίδιο buffer
- υπολογίζει την τυπική απόκλιση¹ για καθεμία από τις αναδιπλώσεις
- αποθηκεύει τη μέγιστη τυπική απόκλιση, την περίοδο στην οποία βρέθηκε (Εξίσωση (1.16)) — χρησιμοποιώντας τα `significance` and `period` στοιχεία της δομής

¹ Στην γλώσσα της επεξεργασίας σήματος είναι η AC ισχύς

- το μέτρο σημαντικότητας *significance* διαιρείται με την τετραγωνική ρίζα του πλήθους των αναδιπλώσεων για να προσομοιώσουμε την αύξηση του λόγου σήματος – θορύβου (βλ. Εξίσωση (1.8))
- η ανάλυση της περιόδου επιστρέφεται ως μέτρο αβεβαιότητας λόγω της διακριτής φύσης της διαδικασίας. Όπως φαίνεται στην Εξίσωση (1.16), είναι ίσο με $\frac{1}{\text{πλήθος αναδιπλώσεων} - 1}$
- η μνήμη απελευθερώνεται και το αρχείο αναφοράς οριστικοποιείται

Επίσης, για τυφλές έρευνες, η παραπάνω διαδικασία ακολουθείται για πολλές περιόδους. Περιορίσαμε την αναζήτηση στο εύρος 8 με $N/8$ στοιχείων. Συνολικά η συνάρτηση *FFAsearch* λαμβάνει τα παρακάτω βήματα:

- δέσμευση μνήμης για το buffer και προετοιμασία αρχείου εξόδου
- η FFA καλείται για όλες τις δοκιμαστικές περιόδους και τα αποτελέσματα (ζευγάρια *period and significance*) αποθηκεύονται το αρχείο της εξόδου
- η μέγιστη τιμή *significance* υποδεικνύει την πιο πιθανή περίοδο
- η μνήμη απελευθερώνεται και αρχείο οριστικοποιείται

Ο κώδικας του FFA παρουσιάζεται στο Παράρτημα Γ'.4

Αναδίπλωση

Η συνάρτηση *FOLD*, αναδιπλώνει την είσοδο σε ένα φασικό διάγραμμα για μία δοσμένη περίοδο που ορίζεται κατά την κλήση της. Ο αλγόριθμος είναι ως εξής:

- προετοιμασία του αρχείου αναφοράς
- ο πρώτος χρόνος παρατήρησης αφαιρείται από κάθε τιμή χρόνου ώστε να καταλήξουμε σε εύρος (0, διάρκεια παρατηρήσεων)

- λαμβάνεται το κλασματικό μέρος της διαίρεσης των χρόνων με τη δοσμένη περίοδο $\frac{t}{p} - \left\lfloor \frac{t}{p} \right\rfloor$ (εύρος: 0 ως 1)
- αφαιρείται 0.5, και καταλήγουμε στη φάση κάθε μέτρησης για τη συγκεκριμένη περίοδο (εύρος: -0.5 ως 0.5)
- η φάσεις και οι κανονικοποιημένες στη μονάδα μετρήσεις αποθηκεύονται στο αρχείο αναφοράς

Το αποτέλεσμα του ανωτέρω αλγορίθμου είναι ένα πίνακας ίσου μεγέθους με τον αρχικό, χωρίς να πραγματοποιείται άθροιση μεταξύ μετρήσεων. Με αυτόν τον τρόπο, μπορούμε εύκολα να διακρίνουμε περιπτώσεις όπου η ανιχνευμένη περίοδος αντιστοιχεί σε συντονισμό δύο περιόδων, αρμονική ή σε τιμή κοντά στην πραγματική περίοδο.

Από την άλλη, αν αναδιπλώσουμε και ταυτόχρονα εξάγουμε μέσους όρους μεταξύ διπλών στοιχείων της χρονοσειράς, τότε σε περίπτωση επιτυχημένης ανίχνευσης, ο θόρυβος θα μειωθεί. Η συνάρτηση FOLD_REBINNED παράγει αυτήν την αναδίπλωση, μόνο που αυτή τη φορά η φάση αποθηκεύεται σε διαστήματα τιμών — phase bins. Τα βήματα που ακολουθεί είναι:

- δεσμεύεται μνήμη για δύο πίνακες: ένας για τα phase bins και ένας για το πλήθος των μετρήσεων που έχουν καταχωρηθεί σε κάθε bin
- προετοιμασία αρχείου αναφοράς
- ο πρώτος χρόνος παρατήρησης αφαιρείται από κάθε τιμή χρόνου ώστε να καταλήξουμε σε εύρος (0, διάρκεια παρατηρήσεων)
- λαμβάνεται το κλασματικό μέρος της διαίρεσης των χρόνων με τη δοσμένη περίοδο $\frac{t}{p} - \left\lfloor \frac{t}{p} \right\rfloor$ (εύρος: 0 ως 1)
- η προηγούμενη τιμή πολλαπλασιάζεται με τον αριθμό των bins και το ακέραιο μέρος λαμβάνεται υπόψη (0 ως BIN-1) δίνοντας τη θέση στον πίνακα των bins στην οποία θα προστεθεί η μέτρηση
- προστίθεται η μέτρηση στο κατάλληλο bin και το πλήθος των μετρήσεων για αυτό, αυξάνεται κατά ένα

- όταν γίνουν τα προηγούμενα βήματα για όλα τα bins, διαιρείται κάθε τιμή με το πλήθος των μετρήσεων που περιλαμβάνει ώστε να λάβουμε τους μέσους όρους, οι οποίοι αποθηκεύονται στο αρχείο αναφοράς
- τα bin με μέσο όρο 0 παραλείπονται στην έξοδο καθώς δεν αντιστοιχεί η τιμή αυτή σε παρατηρήσεις αλλά στην έλλειψή τους

2.2 Τεστ υπολογιστικής πολυπλοκότητας

Με αυτό το τεστ, προσπαθήσαμε να επιβεβαιώσουμε την υπολογιστική πολυπλοκότητα των FFT και FFA. Εδώ απαριθμούμε τα πειραματικά βήματα:

1. Δημιουργία των αρχείων αναφορών `reportffa.txt` και `reportfft.txt`.
2. Μία συνάρτηση δημιουργία τυχαίων καμπύλων φωτός εισήχθη στο λογισμικό. Οι παράμετροι των καμπυλών δεν έχουν σημασία καθώς η ταχύτητα των συγκεκριμένων αλγορίθμων δεν εξαρτάται από το περιεχόμενο αλλά το μέγεθος της εισόδου.
3. Τα μεγέθη της εισόδου ήταν 32, 64, 128, ..., 16777216. Στην περίπτωση του FFA η πολυπλοκότητα $O(n^2)$ μας οδήγησε να λάβουμε όριο τα 262144 στοιχεία.
4. Οι FFT και FFAsearch συναρτήσεις κλήθηκαν σε κάθε περίπτωση και χρονομετρήθηκαν χρησιμοποιώντας CPU clocks.
5. Οι μικρές εισοδοι επεξεργάζονταν μέσα σε χιλιοστά ή και εκατομμυριοστά του δευτερολέπτου, χρόνοι που δε μπορούν να μετρηθούν με ακρίβεια από τις ενσωματωμένες συναρτήσεις της C (ακρίβεια λίγων χιλιοστών του δευτερολέπτου). Για να υπερβούμε αυτό το εμπόδιο, κάθε βήματα επαναλαμβάνονταν τουλάχιστον για 10 δευτερόλεπτα. Με διαίρεση δια του αριθμού των επαναλήψεων, βρήκαμε καλές εκτιμήσεις της διάρκειας εκτέλεσης και των πιο γρήγορων βημάτων.
6. Για καλύτερη στατιστική, κάθε βήμα εκτελούνταν τουλάχιστον 5 φορές, ασχέτως διάρκειας

7. Οι χρόνοι μετατράπηκαν σε ns και αποθηκεύτηκαν στα αρχεία αναφορών. Έπειτα, διαιρέθηκαν με $N \log_2 N$ (FFT) και N^2 (FFA), ώστε να εκτιμηθεί ο συντελεστής του πρώτου όρου της υπολογιστικής πολυπλοκότητας για κάθε μέγεθος εισόδου. Σχετικά σταθεροί συντελεστές θα επιβεβαίωναν την αναμενόμενη συμπεριφορά των αλγορίθμων.

2.3 Τεστ σε κατασκευασμένες καμπύλες φωτός

Η επίδοση των FFA και FFT κάτω από την παρουσία υψηλών επιπέδων θορύβου για διάφορα σήματα, ελέγχθηκε με την κατασκευή καμπύλων φωτός με μεταβαλλόμενες παραμέτρους.

Καθώς η απόλυτη τιμή της ισχύος σήματος ή θορύβου δεν είναι σημαντική, επιλέξαμε να κρατήσουμε σταθερή τη μία — ισχύς σήματος — και να μεταβάλλουμε την άλλη — θόρυβος — σε σχέση με την πρώτη. Επίσης, το μέγεθος της χρονοσειράς δεν επηρεάζει τα αποτελέσματα μετά από ένα ορισμένο όριο: έχουμε σχεδόν συνεχή συμπεριφορά. Χρησιμοποιώντας το γεγονός αυτό, κρατήσαμε σταθερό το μέγεθος για να αποφύγουμε περιττές καθυστερήσεις και μεταβάλαμε την περίοδο του σήματος.

Το σχήμα του σήματος είναι επηρεάζει την πιθανότητα εντοπισμού του. Αν και υπάρχουν πολλές και διάφορες μορφές που συναντά κανείς στο πεδίο της αστρονομίας, επιλέξαμε δύο βασικές που προσεγγίζουν ένα εύρος αστρονομικών φαινομένων: (i) ημιτονοειδής και (ii) αλληλουχία όμοιων παλμών. Η πρώτη προσομοιώνει πολλές μορφές μεταβλητών αστερών. Η δεύτερη μπορεί να αναδείξει τη συμπεριφορά των αλγορίθμων στις περιπτώσεις των pulsar, διαβάσεων εξωηλιακών πλανητών κτλ. — όπου στενοί παλμοί μας ενδιαφέρουν.

Ωστόσο, οι παλμοί χρειάζονται μία ακόμη παράμετρο για να οριστούν: τον κύκλο εργασίας (D). Αυτός, είναι η διάρκεια ενός ‘γεγονότος’ που μας ενδιαφέρει προς την περίοδο του σήματος. Για παράδειγμα, στην περίπτωση των εξωηλιακών πλανητών που εντοπίζονται με τη μέθοδο των διαβάσεων, η διάρκεια της διάβασης (π.χ. λίγες ώρες) προς την τροχιακή περίοδο (π.χ. λίγες μέρες ή βδομάδες) του πλανήτη δίνει τον κύκλο εργασίας.

Η φάση των περιοδικών σημάτων θεωρείται ότι δεν αλλοιώνει τα αποτελέσματα των αλγορίθμων οπότε και επιλεγόταν τυχαία για κάθε κατασκευασμένη καμπύλη.

Περιγραφή του πειράματος

1. Μία συνάρτηση παράγει τεχνητές καμπύλες φωτός. Οι παράμετροι είναι:
 - DC offset
 - επίπεδο θορύβου
 - (i) πλάτος, (ii) περίοδος και (iii) φάση ημιτονοειδούς συνιστώσας
 - (i) πλάτος, (ii) περίοδος, (iii) φάση και (iv) κύκλος εργασίας για τους ορθογώνιους παλμούς
2. Το DC offset τίθεται 0 καθώς οι αλγόριθμοι δεν εξαρτώνται από αυτό — υπολογίζουν τυπικές αποκλίσεις και όχι ισχύ
3. Το επίπεδο θορύβου επιλέγεται στο εύρος 0.0 to 3.0 φορές του πλάτους του σήματος το οποίο είναι σταθερό (1.0) και παίρνει 20 διαδοχικές τιμές
4. Η περίοδος έπαιρνε 50 διαφορετικές τιμές στο εύρος 32 ως 128 στοιχείων στη χρονοσειρά των 1024 στοιχείων.
5. Η φάση επιλεγόταν τυχαία για κάθε καμπύλη φωτός
6. Τα τεστ έγιναν για 4 σχήματα:
 - Ημιτονοειδές
 - Ορθογώνιο κύκλου εργασίας 20%
 - Ορθογώνιο κύκλου εργασίας 10%
 - Ορθογώνιο κύκλου εργασίας 5%
7. Γνωρίζοντας τη σωστή περίοδο, καταγράψαμε τα σφάλματα των εκτιμήσεων των αλγορίθμων FFT και FFA. Επιλέξαμε ως επίπεδο ανοχής το 10% για το χαρακτηρισμός μίας ανίχνευσης ως επιτυχημένης. Σε αντίθετη περίπτωση το σφάλμα λάμβανε νέα τιμή, το 1 (100% αποτυχία). Όπως θα φανεί

αργότερα, αυτό θα βοηθήσει στην κατανόηση της συμπεριφοράς των αλγορίθμων μέσα από διαγράμματα.

8. Μία σειρά τριδιάστατων διαγραμμάτων παράχθηκε για να αναδείξει την εξάρτηση της ανίχνευσης από την περίοδο και το επίπεδο θορύβου για όλες τις 8000 καμπύλες φωτός που δημιουργήθηκαν.

2.4 Τεστ σε δεδομένα της αποστολής Kepler

Καθώς οι αλγόριθμοι FFA και FFT, έχουν ως είσοδο τακτικές χρονοσειρές, κανείς πρέπει να είναι προσεκτικός την επιλογή αστρονομικών δεδομένων. Για τους λόγους που αναλύθηκαν στην §1.1 επιλέξαμε παρατηρησιακά δεδομένα που πάρθηκαν από το διαστημικό τηλεσκόπιο Kepler.

Ένας άλλος λόγος που επιλέξαμε αυτήν την αποστολή ήταν η διαθεσιμότητα των δεδομένων καθώς και πολλών δημοσιεύσεων με αποτελέσματα που μπορούμε να συγκρίνουμε με τα δικά μας.

2.4.1 Επιλογή των στόχων

Η ερευνητική ομάδα του προγράμματος All Sky Automated Survey (ASAS) έχει δημοσιεύσει μία λίστα 947 μεταβλητών αστερών στο πεδίο του Kepler (Pigulski et al., 2009). Ο τύπος καθώς και η περίοδος του καθενός δίνονται στο ίδιο paper. Χρησιμοποιώντας την έρευνα αυτή ως οδηγό, επιλέξαμε τέσσερα αντικείμενα από τις εξής κατηγορίες:

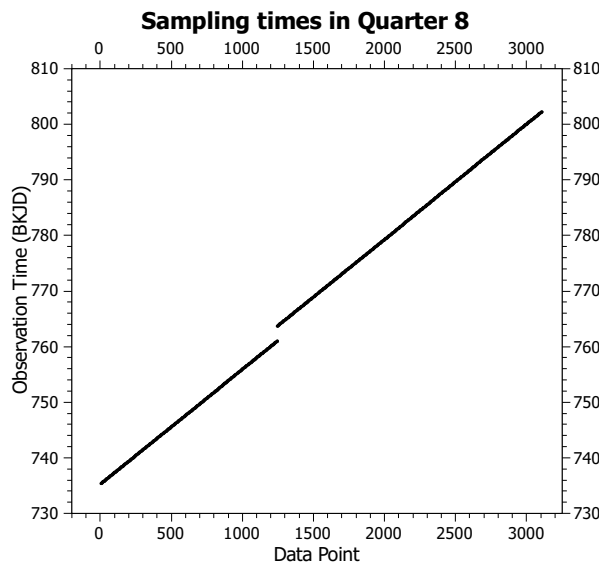
- Κηφείδες μεταβλητοί
- Διπλά εκλειπτικά συστήματα
- Αστέρες με διαβάσεις εξωπλανητών

2.4.2 Διαχείριση καμπυλών φωτός

Το Kepler ID του κάθε στόχου του ASAS αναγνωρίστηκε μέσα από τις συνεταγμένες στο ουρανογραφικό σύστημα αναφοράς (ορθή αναφορά και απόκλιση). Έπειτα, οι καμπύλες φωτός των αστερών λήφθηκαν από το *NASA Exoplanet Archive*¹. Τα αρχεία περιείχαν τις ακριβείς μετρήσεις χρόνου για κάθε παρατήρηση σε BJD.

2.4.3 Τακτοποίηση χρονοσειρών σχεδόν ίσων διαστημάτων

Παρά το ότι η συχνότητα δειγματοληψίας του διαστημικού τηλεσκοπίου Kepler είναι πολύ σταθερή (περίπου 30 λεπτά για το long cadence και 1 λεπτό για το short cadence), αποφασίσαμε να ελέγξουμε αυτήν την πληροφορία.

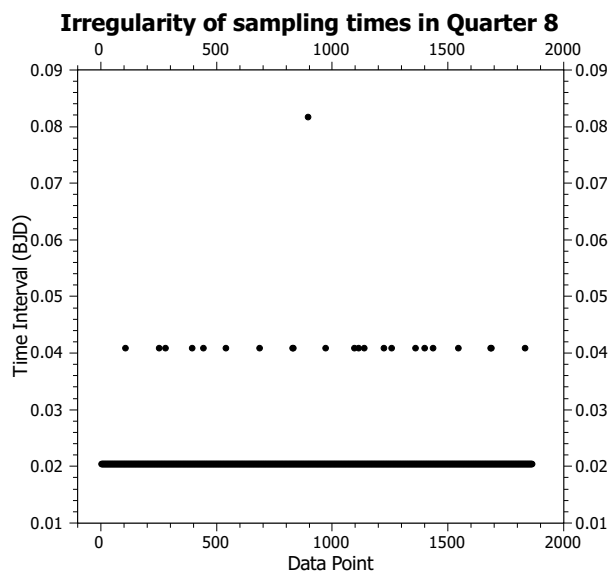


Σχήμα 2.1: Χρόνοι δειγματοληψίας στο τρίμηνο 8 της αποστολής Kepler. Σημειώστε το κενό τριών ημερών στο δεύτερο μήνα.

Στα περισσότερα τρίμηνα παρατηρήσεων της αποστολής Kepler, υπάρχουν κενά λίγων ημερών που αντιστοιχούν στα Safe Modes του διαστημικού τηλεσκοπίου για τεχνικούς ή προληπτικούς λόγους. Για παράδειγμα, στο τρίμηνο 8 υπήρξαν τρία Safe Modes. Το πρώτο τέθηκε στο μεσοδιάστημα των τριμήνων 7 και 8,

¹ http://exoplanetarchive.ipac.caltech.edu/applications/ETSS/Kepler_index.html

προκαλώντας μία καθυστέρηση στη συνέχιση των παρατηρήσεων. Το δεύτερο, συνέβη στο δεύτερο μήνα του τριμήνου. Το τρίτο λαμβάνοντας χώρα στο τέλος του τριμήνου διέκοψε πρόωρα τις παρατηρήσεις (Σχήμα 2.1) (Christiansen et al., 2011).



Σχήμα 2.2: Ασυνέχειες στη δειγματοληψία του Kepler στο 8ο τρίμηνο παρατηρήσεων. Σημειώστε ότι δεν αναφερόμαστε σε Safe Modes του τηλεσκοπίου που δημιουργούν πολύ μεγαλύτερα κενά στα δεδομένα.

Τώρα, έχοντας αφαιρέσει τα δεδομένα πριν το κενό, εστιάζουμε στις παρατηρήσεις που εμφανίζονται συνεχείς. Με τον υπολογισμό διαδοχικών διαφορών στους χρόνους παρατήρησης καταλήγουμε σε μία εκτίμηση της περιόδου δειγματοληψίας σε κάθε σημείο δεδομένων. Για μία τακτική χρονοσειρά το διάγραμμα της περιόδου δειγματοληψίας είναι μία ευθεία γραμμή. Αντιθέτως, παρατηρήσαμε έναν αριθμό περιπτώσεων με κενά λίγων samples (Σχήμα 2.2). Αυτές οι αποκλίσεις αντιστοιχούν σε μετρήσεις που δε λήφθηκαν υπόψη λόγω momentum desaturations: περίπου κάθε 3 μέρες οι προωθητήρες του τηλεσκοπίου διορθώνουν ανεπιθύμητες μεταβολές της γωνιακής του ταχύτητας που προκαλούνται από την πίεση της ηλιακής ακτινοβολίας (εφαρμογή ροπής) (Christiansen et al., 2011).

Αν αφαιρέσουμε τα ‘ελαττωματικά’ σημεία, η περίοδος δειγματοληψίας έχει (i) μέση τιμή 0.0204341 BJD (29 λεπτά και 25.5 δευτερόλεπτα) και (ii) τυπική απόκλιση $1.489 \cdot 10^{-7}$ BJD (13 ms). Αυτή η μισάωρη δειγματοληψία αντιστοιχεί στη

‘long cadence’ ρύθμιση. Ωστόσο, τα διαστήματα μεταξύ momentum desaturations είναι μόλις λίγες μέρες και μία αποκοπή των δεδομένων (όπως στην περίπτωση των Safe Modes) δεν είναι επιθυμητή.

Ο μέσος όρος και η τυπική απόκλιση των διαστημάτων ανάμεσα στις παρατηρήσεις χωρίς να απορριφθούν τα ‘κακά’ σημεία, είναι 0.0207194 BJD (29 λεπτά και 50.2 δευτερόλεπτα) και 0.00266 BJD (3 λεπτά και 50.2 δευτερόλεπτα) αντίστοιχα. Όπως βλέπουμε, τα λίγα αυτά σημεία μπορούν να οδηγήσουν σε μικρές ασάφειες σε εκτιμήσεις περιόδων και χρόνων. Εξάλλου, ο κώδικας θα πρέπει να μπορεί να διαχειριστεί μετρήσεις και από έρευνες με όχι τόσο καλή δειγματοληψία όσο του Kepler.

Προτεινόμενη λύση

Ένας τρόπος να υπερβούμε τη δυσκολία αυτή¹ είναι να εφαρμόσουμε τους χρόνους παρατήρησης σε ένα γραμμικό μοντέλο με τη μέθοδο των ελαχίστων τετραγώνων (LSM) (παρόλο που υποθέτει Γκαουσιανή κατανομή των υπολοίπων, Wall & Jenkins (2003)):

- Λαμβάνουμε τη λίστα των N χρόνων παρατήρησης t_i , όπου $i \in \{0, 1, \dots, N-1\}$.
- Θεωρούμε το γραμμικό μοντέλο που εξαρτάται από τα (i) τ_0 – αρχή παρατήρησης και (ii) p_s περίοδος δειγματοληψίας. Οι N μετρήσεις αντιστοιχούν σε χρόνους τ_i όπου $\tau_i = \tau_0 + i \cdot p_s$.
- Ως μέτρο καλής προσαρμογής επιλέγουμε το άθροισμα των τετραγώνων των υπολοίπων: $g = \sum_{i=0}^{N-1} (t_i - \tau_i)^2$ που προφανώς παρουσιάζει ελάχιστον για την καλύτερη προσαρμογή
- Υπολογίζουμε τις τιμές των τ_0 and p_s για βέλτιστη προσαρμογή λύνοντας το σύστημα:
$$\begin{cases} \frac{\partial g}{\partial \tau_0} = 0 \\ \frac{\partial g}{\partial p_s} = 0 \end{cases}.$$

¹ υπάρχουν μέθοδοι που εφευρέθηκαν μετά τους FFT και FFA για ανάλυση ανομοιογενών χρονοσειρών αλλά δε μπορούν να αναλυθούν στην παρούσα εργασία. Για περισσότερες πληροφορίες, βλέπε περιοδόγραμμα Lomb–Scargle: Lomb (1976); Scargle (1982).

- Αποδεικνύεται (βλ. § Α'3) ότι:

$$\tau_0 = \frac{2}{N^2 + N} \left[(2N - 1) \sum_{i=0}^{N-1} t_i - 3 \sum_{i=0}^{N-1} (i \cdot t_i) \right] \quad (2.1)$$

$$p_s = \frac{6}{N(N^2 - 1)} \left[2 \sum_{i=0}^{N-1} (i \cdot t_i) - (N - 1) \sum_{i=0}^{N-1} t_i \right] \quad (2.2)$$

- Από την τιμή του μέτρου καλής προσαρμογής για τις τιμές που υπολογίστηκαν (g_{min}), προκύπτει μία τιμή σφάλματος (regression error) το οποίο θα αποκαλούμε 'σφάλμα διακριτοποίησης χρόνου' και θα συμβολίζουμε ως t_{err} . Από τον ορισμό του g και της τυπικής απόκλισης, και την επιλογή μας για χρήση 2.58 σ (99% των τιμών μέσα στα όρια αυτά, υποθέτοντας κανονική κατανομή), ορίζουμε:

$$t_{err} = 2.58 \sqrt{\frac{g_{min}}{N - 2}} \quad (2.3)$$

Σημείωση: η διαίρεση με $N - 2$ υπαγορεύεται από το διπαραματρικό μοντέλο.

Μετά από αυτήν τη διαδικασία, οποιαδήποτε επόμενη ανάλυση μπορεί να κάνει χρήση της νέας περιόδου δειγματοληψίας και να λάβει υπόψη το σφάλμα διακριτοποίησης. Για τον κώδικα, παραπέμπουμε στο Παράρτημα Γ'2.

2.4.4 Ανάλυση των δεδομένων

Η ανάλυση της καμπύλης φωτός καθενός από τους στόχους περιελάμβανε τα εξής βήματα:

- Ανάγνωση του αρχείου του αστέρα και καταγραφή των μετρήσεων και των χρόνων τους
- Τακτοποίηση της χρονοσειράς
- Εφαρμογή των FFT και FFA. Αποθήκευση των εκτιμήσεων περιόδων

- Παραγωγή αναδιπλώσεων re-binned και μη, και για τις δύο εκτιμήσεις περιόδων

“Αποτελέσματα; Γιατί, έχω μπόλικά από αυτά! Αν βρω 10000 τρόπους που κάτι δε λειτουργεί, δεν απέτυχα. Δεν αποθαρρύνομαι, γιατί κάθε λανθασμένη προσέγγιση είναι συχνά ένα βήμα μπρος...”

Thomas Alva Edison (Αμερικανός εφευρέτης)

3

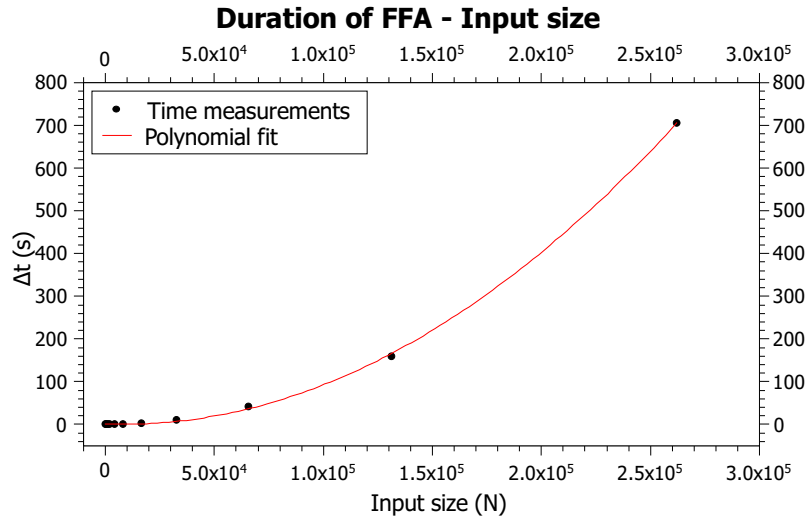
Αποτελέσματα

3.1 Τεστ υπολογιστικής πολυπλοκότητας

Σε αυτήν την ενότητα βρίσκονται τα αποτελέσματα του τεστ της ταχύτητα εκτέλεσης των αλγορίθμων σε σχέση με το μέγεθος της εισόδου με την οποία τροφοδοτούνται. Στις επόμενες σελίδες παρουσιάζονται διαγράμματα Διάρκειας – Μεγέθους Εισόδου καθώς και των συντελεστών των σημαντικότερων όρων της αναμενόμενης υπολογιστικής πολυπλοκότητας ($n \log n$ για τον FFT και n^2 για τον ‘τυφλής αναζήτησης’ FFA).

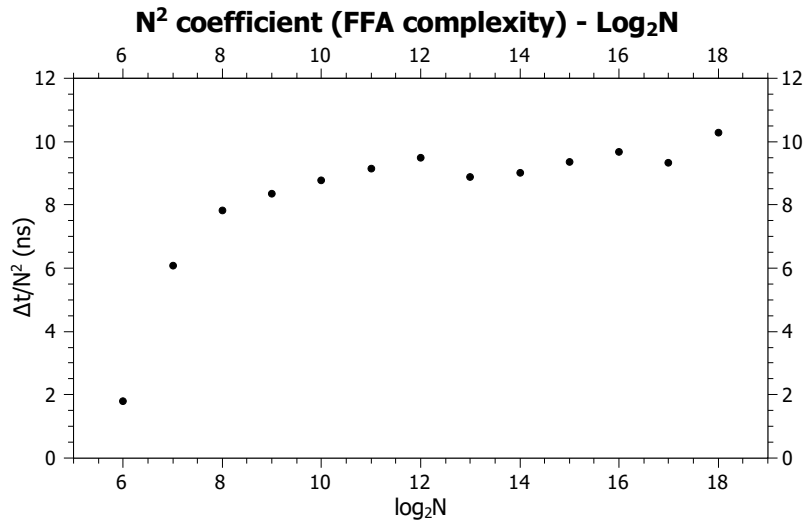
Πίνακας 3.1: Αποτελέσματα τεστ υπολογιστικής πολυπλοκότητας

Μέγεθος Εισόδου	FFT		FFA	
	Χρονος (s)	Σταθερά του $N \log_2 N$ (ns)	Χρόνος (s)	Σταθερά του N^2 (ns)
32	0.000013	79.431076	0.000001	0.046539
64	0.000027	70.500275	0.000007	1.802245
128	0.000059	65.380890	0.000100	6.087590
256	0.000132	64.414253	0.000513	7.825421
512	0.000294	63.883896	0.002188	8.346994
1024	0.000699	68.224289	0.009205	8.778537
2048	0.001462	64.917915	0.038372	9.148514
4096	0.003310	67.334999	0.159222	9.490384
8192	0.006983	65.572792	0.596471	8.888104
16384	0.014793	64.493263	2.418000	9.007751
32768	0.034235	69.652298	10.043400	9.353645
65536	0.119595	114.054901	41.555199	9.675323
131072	0.271108	121.670052	160.449402	9.339384
262144	0.565944	119.939262	706.730835	10.284287
524288	1.229000	123.375336		
1048576	2.427400	115.747459		
2097152	5.070000	115.122116		
4194304	10.667400	115.604836		
8388608	23.297001	120.748650		
16777216	59.145802	146.890182		

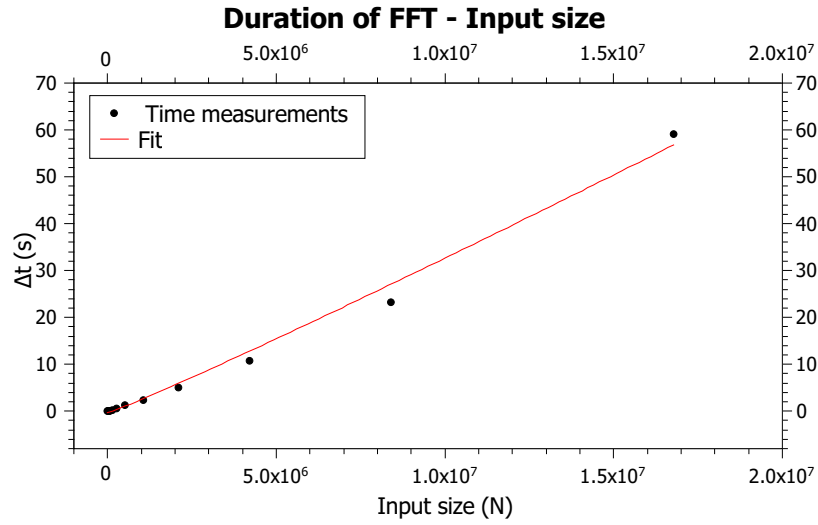


Σχήμα 3.1: Εξάρτηση χρόνου εκτέλεσης του FFA από το μέγεθος της εισόδου

Η εξίσωση της προσαρμογής είναι $0.96 - 0.17 \cdot 10^{-4}N + 1.09 \cdot 10^{-8}N^2$ με $R^2 = 0.99985$.

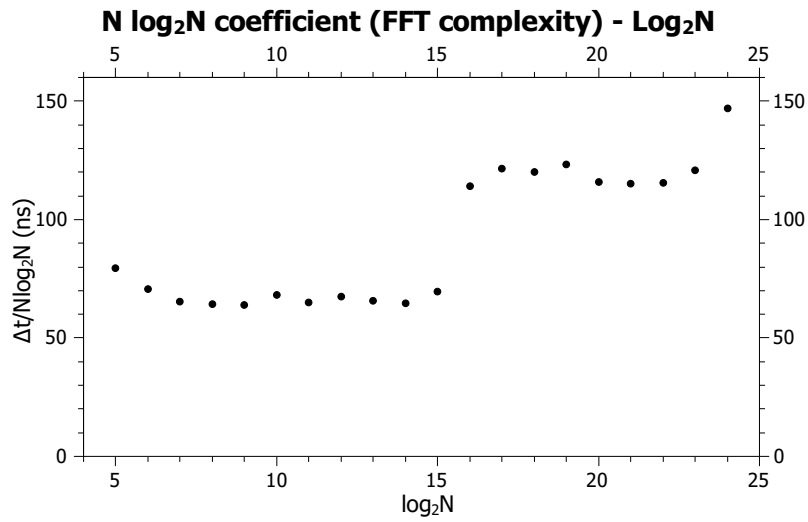


Σχήμα 3.2: Αποτελέσματα πολυπλοκότητας FFA



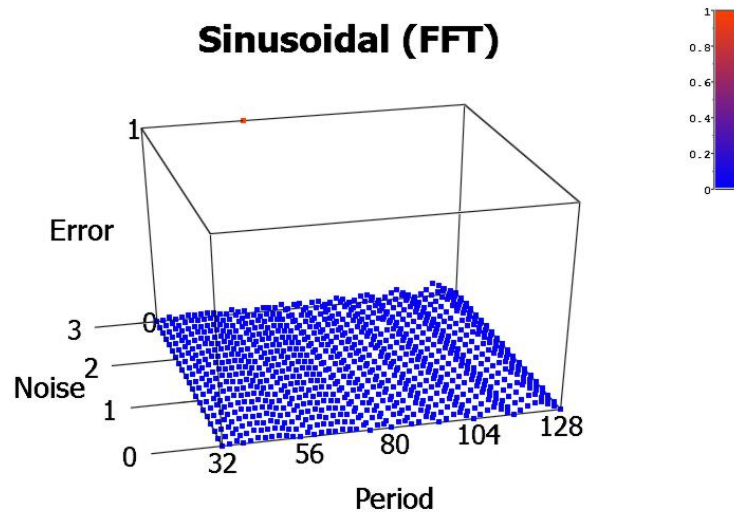
Σχήμα 3.3: Εξάρτηση χρόνου εκτέλεσης του FFT από το μέγεθος της εισόδου

Η εξίσωση της προσαρμογής είναι $-0.33 + 1.42 \cdot 10^{-7} N \log_2 N$ με $R^2 = 0.99287$.

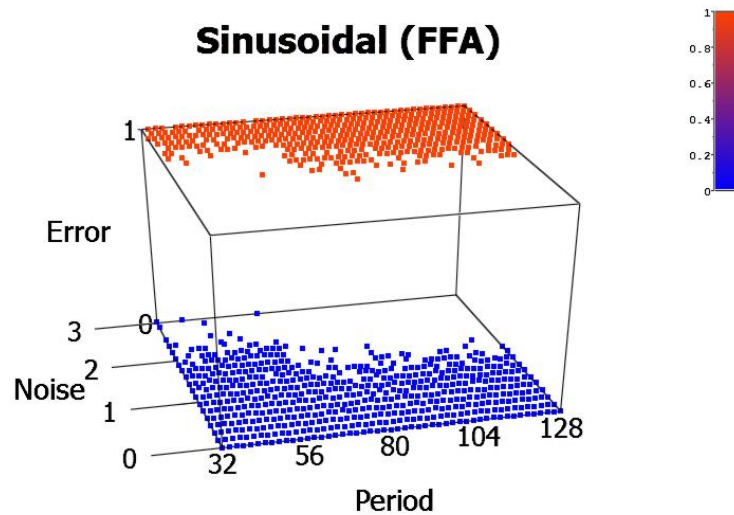


Σχήμα 3.4: Αποτελέσματα πολυπλοκότητας FFT

3.2 Τεστ σε κατασκευασμένες καμπύλες φωτός

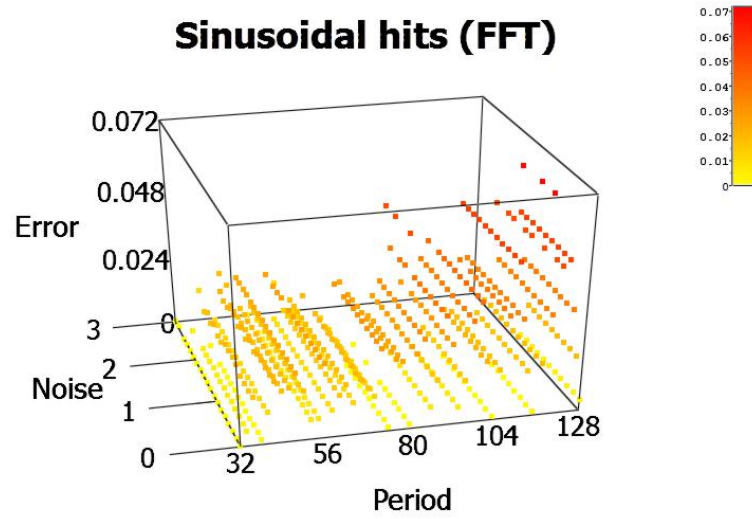


(α') Επιτυχίες και αποτυχίες από τον FFT

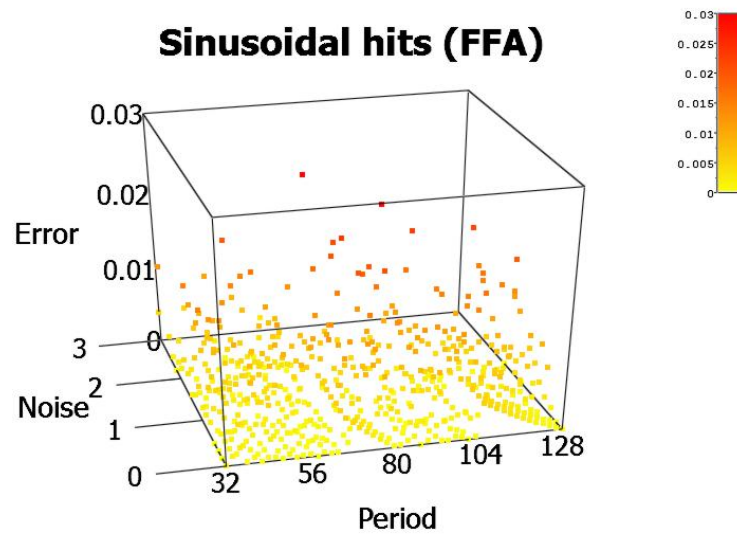


(β') Επιτυχίες και αποτυχίες από τον FFA

Σχήμα 3.5: Επιτυχία ανίχνευσης ημιτονοειδών σημάτων για διάφορες περιόδους και επίπεδα θορύβου

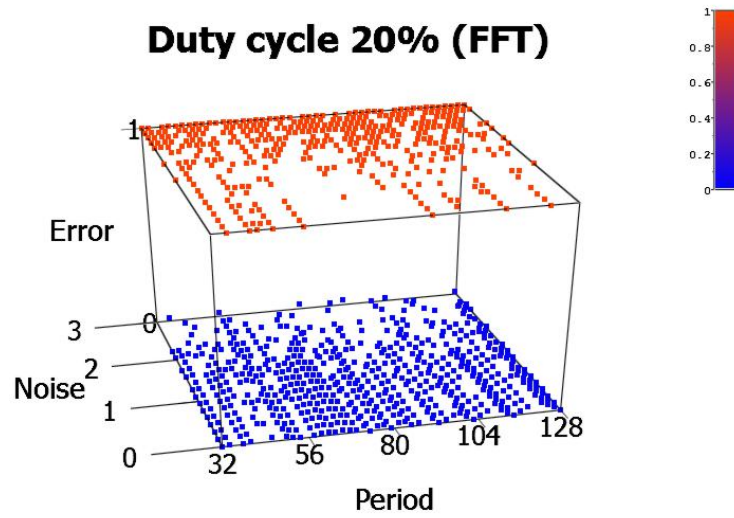


(α') Επιτυχίες με FFT

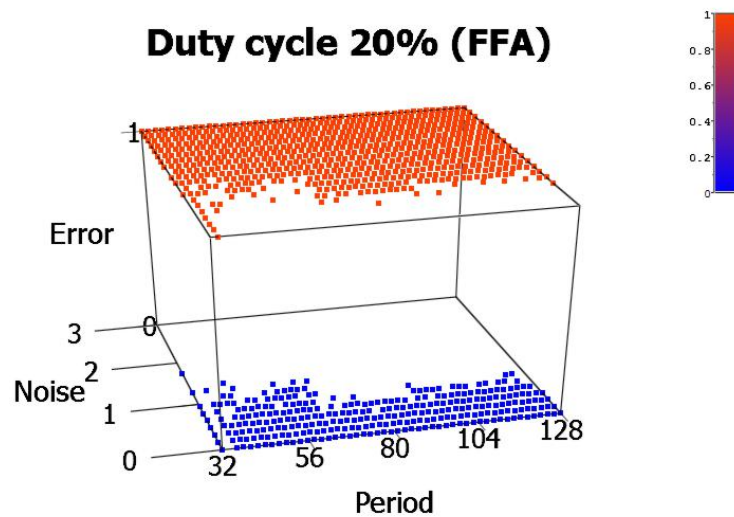


(β') Επιτυχίες με FFA

Σχήμα 3.6: Σφάλματα εντοπισμού ημιτονοειδών σημάτων

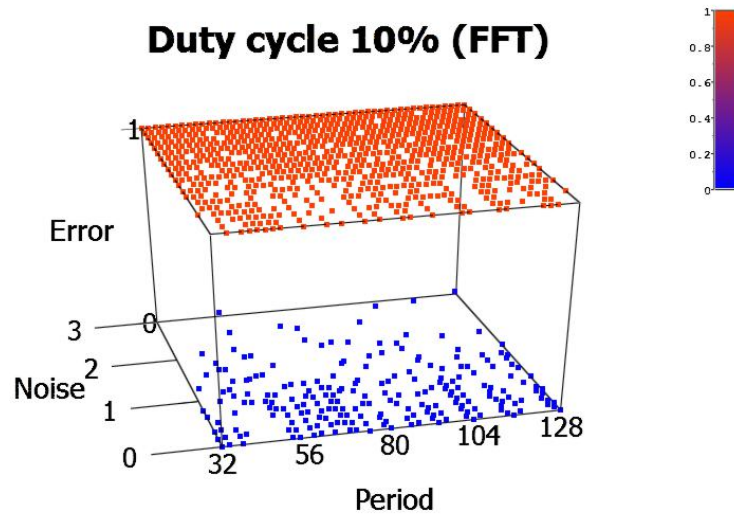


(α') Επιτυχίες και αποτυχίες από τον FFT

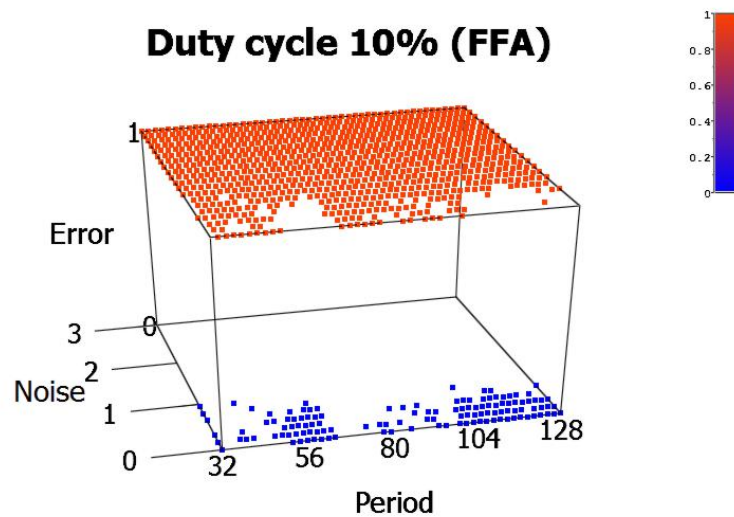


(β') Επιτυχίες και αποτυχίες από τον FFA

Σχήμα 3.7: Επιτυχία ανίχνευσης ορθογώνιων σημάτων κύκλου εργασίας 20% για διάφορες περιόδους και επίπεδα θορύβου

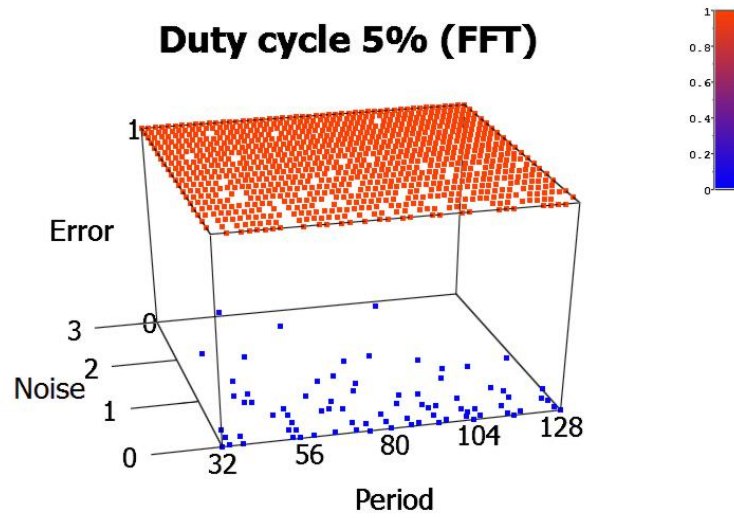


(α') Επιτυχίες και αποτυχίες από τον FFT

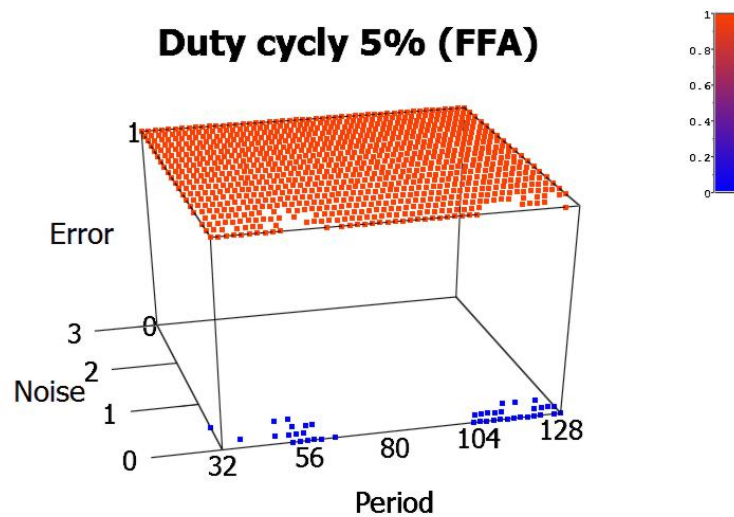


(β') Επιτυχίες και αποτυχίες από τον FFA

Σχήμα 3.8: Επιτυχία ανίχνευσης ορθογώνιων σημάτων κύκλου εργασίας 10% για διάφορες περιόδους και επίπεδα θορύβου



(α') Επιτυχίες και αποτυχίες από τον FFT



(β') Επιτυχίες και αποτυχίες από τον FFA

Σχήμα 3.9: Επιτυχία ανίχνευσης ορθογώνιων σημάτων κύκλου εργασίας 5% για διάφορες περιόδους και επίπεδα θορύβου

3.3 Ανάλυση δεδομένων του Kepler

3.3.1 Παράδειγμα εξόδου προγράμματος

```

1 INPUT REPORT
2
3 Kepler ID          7548061
4 Time offset        2454833.000000 BJD
5 ASAS period        4.926000 days
6 Notes              Q14s
7 Bins                49957
8
9 REGULARIZE REPORT
10
11 New time offset    2456170.169175 BJD
12 Sampling rate      0.984465 min
13 Regularization error 9.043122 min
14
15 FFT REPORT
16
17 > Spectrum file created.
18
19 Frequency:          5 bins
20 Period:              4.525666 days   +/- 0.452567
21
22 > Folding at p=4.525666 saved to: foldfft.txt
23
24 > Re-binned (256) folding at p=4.525666 saved to: foldfftr.txt
25
26 FFA REPORT
27
28 > Periodogram file created.
29
30 Period:              4.931896 days   +/- 0.000114
31
32 > Folding at p=4.931896 saved to: foldffa.txt
33
34 > Re-binned (256) folding at p=4.931896 saved to: foldffar.txt
35
36 Press any key to continue . . .

```

Listing 3.1: Η έξοδος του προγράμματος μετά την ανάλυση της καμπύλης φωτός ενός αστέρα με δεδομένα short cadence του Kepler

3.3.2 Αποτελέσματα

Εδώ παρουσιάζεται μία λίστα με όλους τους αστέρες του πεδίου του Kepler που αναλύθηκαν και την περίοδο που ανίχνευσαν οι αλγόριθμοι FFA και FFT. Επίσης δίνεται η απόκλιση από τη δημοσιευμένη τιμή (Pigulski et al., 2009; O'Donovan et al., 2006; Welsh et al., 2010; Shporer et al., 2011; Borucki et al., 2011) (στη στήλη *Περίοδος* στον τομέα *Περιγραφή στόχου*) ως σφάλμα επί τις εκατό. Σημείωση: οι περίοδος που ανιχνεύθηκε συχνά πολλαπλασιάστηκε με έναν παράγοντα 2 ή 3 για περιπτώσεις εσφαλμένης αρμονικής (βλ. Συζήτηση).

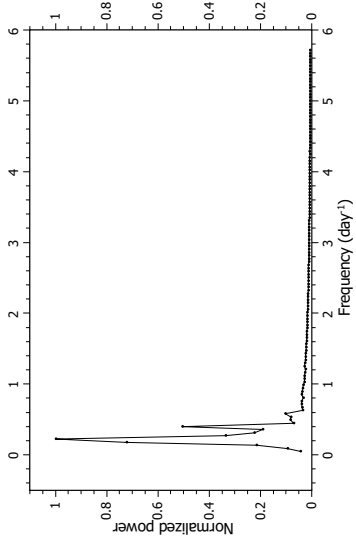
Πίνακας 3.2: Σύνοψη αποτελεσμάτων στα δεδομένα του Kepler

Kepler ID	Περιγραφή στόχου			FFT		FFA	
	RA (° (J2000.0)	Dec (°)	Περίοδος (d)	Περίοδος (d)	Err (%)	Περίοδος (d)	Err (%)
Κηφείδες μεταβλητοί							
7548061	297.064	43.127	4.9260	4.5267	8.11	4.9319	0.12
7899428	294.650	43.692	11.573	11.029	4.70	11.471	0.88
11347875	290.467	49.199	3.4521	3.3454	3.09	3.4322	0.58
12406908	290.937	51.270	13.367	14.883	11.3	3.4322	0.37
Διπλοί εκλειπτικοί							
4660997	293.514	39.711	0.56256	0.56669	0.73	0.56177	0.14
4544587	285.886	39.683	2.1893	2.1745	0.68	2.1913	0.09
8868650	284.398	45.129	4.4472	4.3453	2.29	4.4460	0.03
10618251	298.125	47.812	0.43742	0.43815	0.17	0.43693	0.11
Αστέρες με διαβαίνοντες εξωπλανήτες							
11446443	286.808	49.316	2.4706	2.5104	1.61	0.0034	99.9
10666592	292.247	47.970	2.2047	2.2444	1.80	0.0035	99.8
9941662	286.971	46.868	1.7637	1.7689	0.29	0.0027	99.8
11414511	297.015	49.225	2.8165	2.8110	0.20	0.0030	99.9

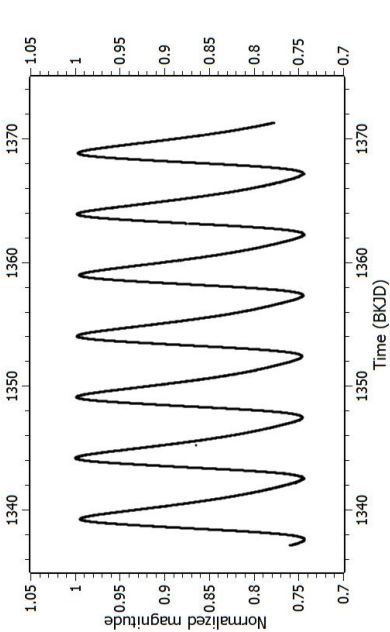
Στη συνέχεια ακολουθούν για κάθε έναν από τους δώδεκα στόχους: πίνακας χαρακτηριστικών, καμπύλη φωτός, φάσμα FFT, περιόδγραμμα FFA, αναδίπλωση re-binned και μη, και στις δύο περιόδους που ανιχνεύονται. Η εκτενής ανάλυση στοχεύει στην καλύτερη απεικόνιση των διαδικασιών που πραγματευόμαστε και στην επίβλεψη της ποιότητας των αποτελεσμάτων.

Kepler ID	7548061
Offset χρόνου	2454833 BJD
Τρίμηνο	14 (short)
Τύπος (ASAS)	Κηφείδης
Περίοδος (ASAS)	4.9260
Μετρήσεις	49957
Περίοδος FFT	4.526 660(452567)
Περίοδος FFA	4.931 896(114)

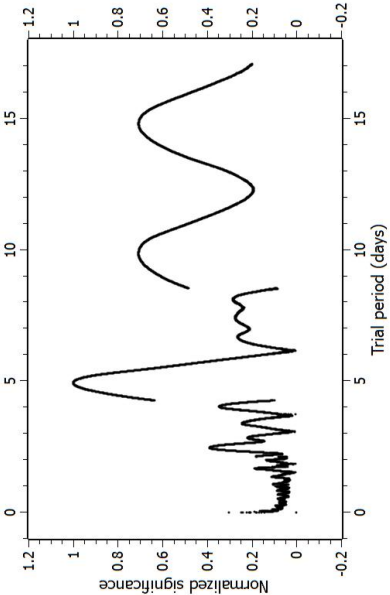
(α') Περιγραφή & ανιχνευμένες περιόδοι



(γ') Φάσμα FFT

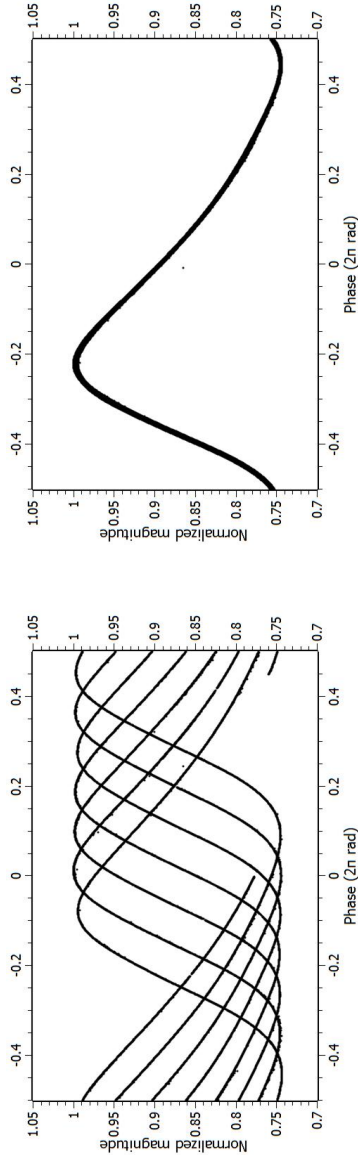


(β') Καμπύλη φωτός



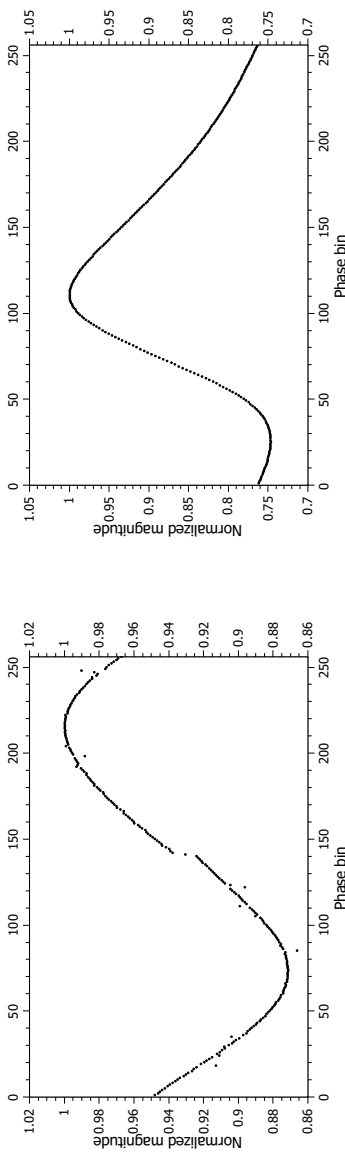
(δ') Περιοδογράμια FFA

Σχήμα 3.10: Αποτελέσματα (α) για KID 7548061 — HIP 97439



(β') Αναδίπλωση στην περίοδο του FFA

(α') Αναδίπλωση στην περίοδο του FFT



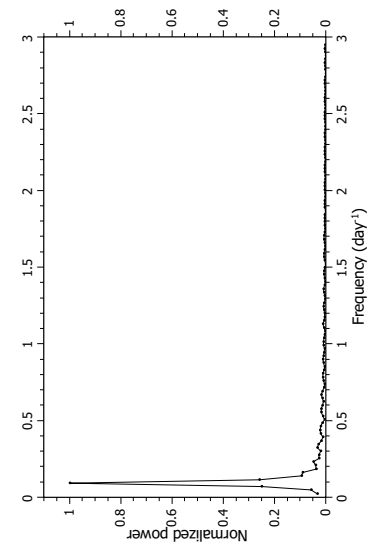
(γ') Re-binned αναδίπλωση στην περίοδο του FFT

(δ') Re-binned αναδίπλωση στην περίοδο του FFA

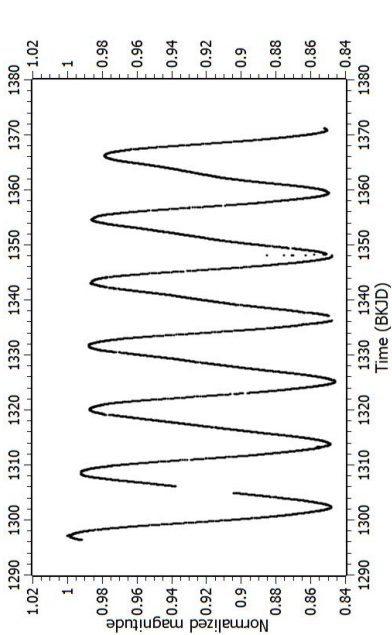
Σχήμα 3.11: Αποτελέσματα (b) για KID 7548061 — HIP 97439

Kepler ID	7899428
Offset χρόνου	2454833 BJD
Τρίμηνο	14 (long)
Τύπος (ASAS)	Κηφείδης
Περίοδος (ASAS)	11.573
Μετρήσεις	3535
Περίοδος FFT	11.028 791 (1378599)
Περίοδος FFA	11.471 383 (3534)

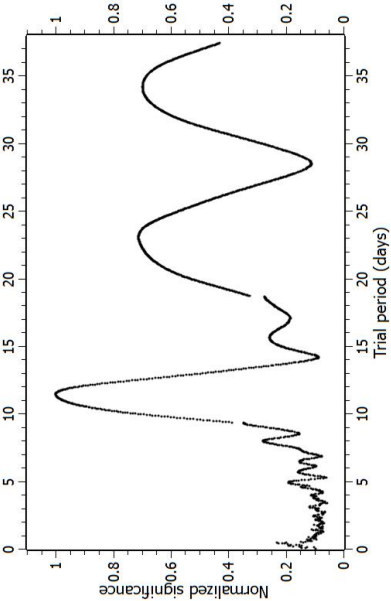
(α') Περιγραφή & ανιχνευμένες περιόδοι



(γ') Φάσμα FFT

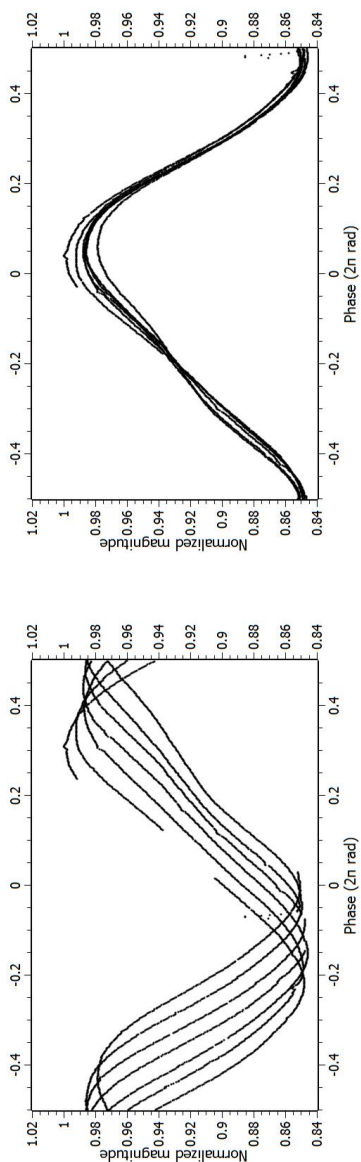


(β') Καμπύλη φωτός



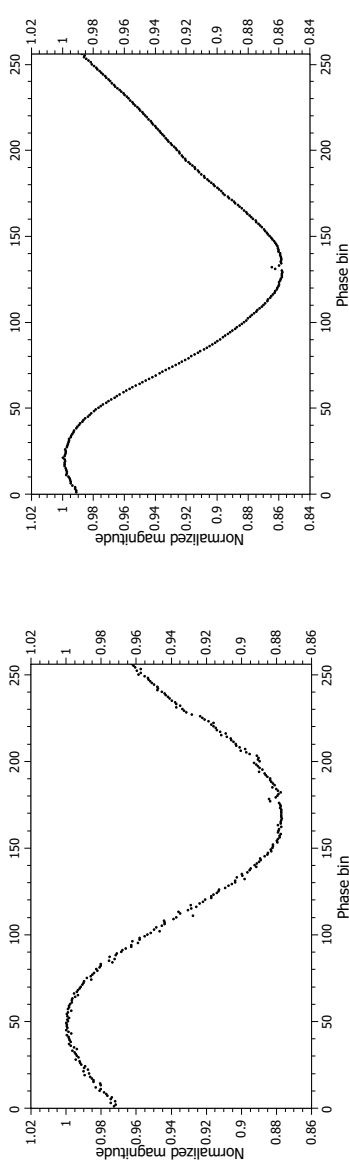
(δ') Περιοδογράμμο FFA

Σχήμα 3.12: Αποτελέσματα (α) για KID 7899428 — TYC 3147–400–1



(β') Αναδίπλωση στην περίοδο του FFA

(α') Αναδίπλωση στην περίοδο του FFT

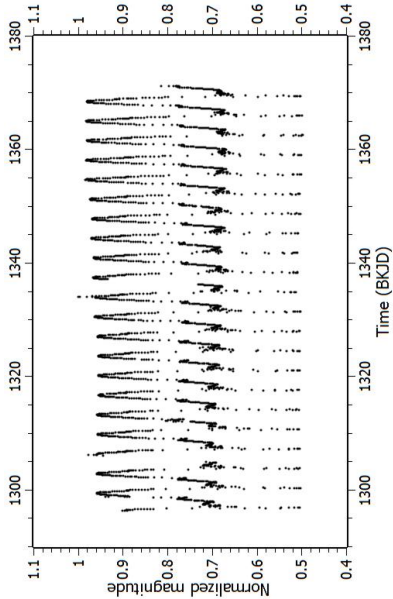


(δ') Re-binned αναδίπλωση στην περίοδο του FFA

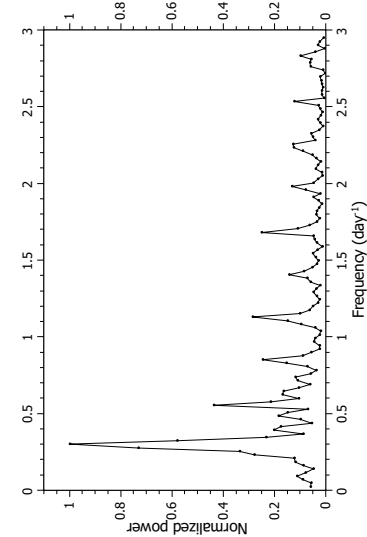
(γ') Re-binned αναδίπλωση στην περίοδο του FFT

Σχήμα 3.13: Αποτελέσματα (b) για KID 7899428 — TYC 3147-400-1

Kepler ID	11347875
Offset χρόνου	2454833 BJD
Τρίμηνο	14 (long)
Τύπος (ASAS)	Κηφείδης
Περίοδος (ASAS)	3.4521
Μετρήσεις	3535
Περίοδος FFT	3.345 397 (128669)
Περίοδος FFA	3.432 224 (707)

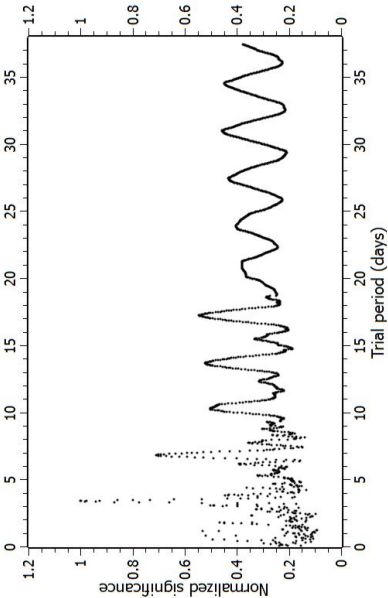


(α') Περιγραφή & ανιχνευμένες περιόδοι



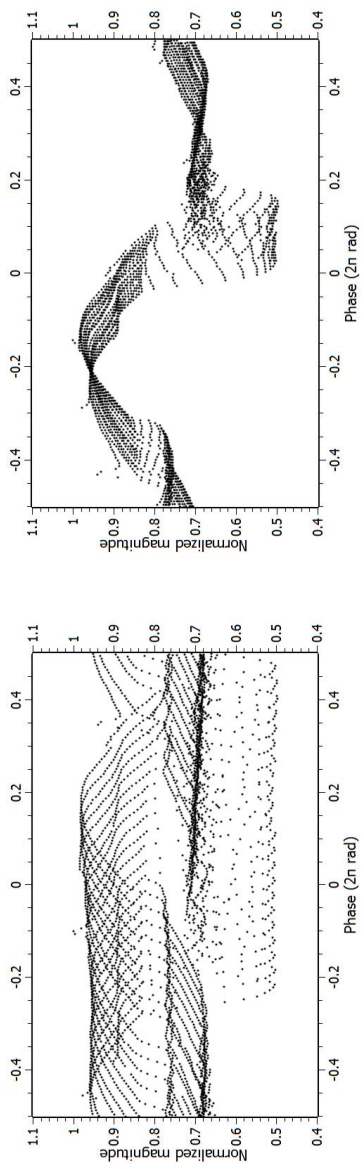
(γ') Φάσμα FFT

(β') Καμπύλη φωτός



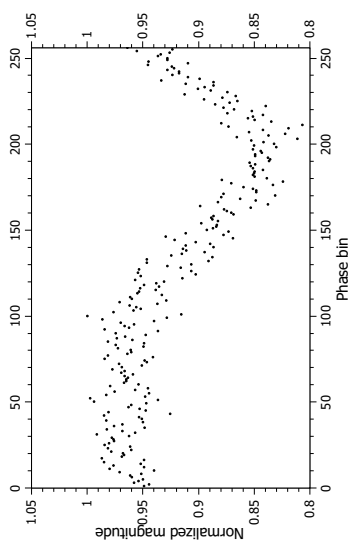
(δ') Περιοδογράμμο FFA

Σχήμα 3.14: Αποτελέσματα (α) για KID 11347875 —2MASS J19215214+4911551

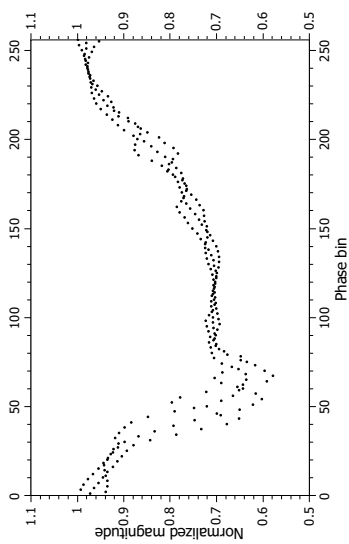


(α') Αναδίπλωση στην περίοδο του FFT

(β') Αναδίπλωση στην περίοδο του FFA



(γ') Re-binned αναδίπλωση στην περίοδο του FFT

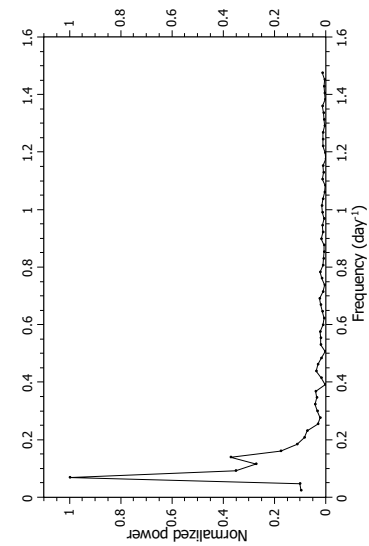


(δ') Re-binned αναδίπλωση στην περίοδο του FFA

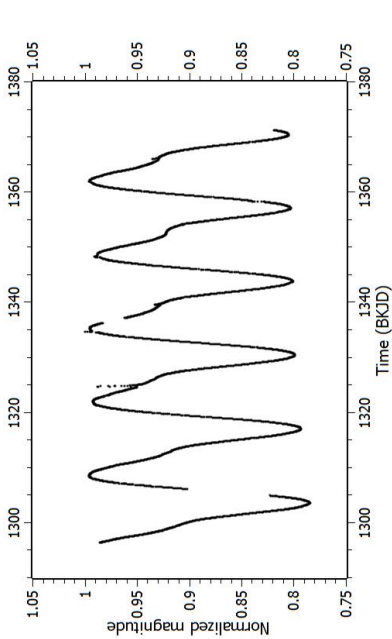
Σχήμα 3.15: Αποτελέσματα (b) για KID 11347875 — 2MASS J19215214+4911551

Kepler ID	12406908
Offset χρόνου	2454833 BJd
Τρίμηνο	14 (long)
Τύπος (ASAS)	Κηφείδης
Περίοδος (ASAS)	13.367
Μετρήσεις	3536
Περίοδος FFT	14.883 124(2480521)
Περίοδος FFA	13.318 062(3533)

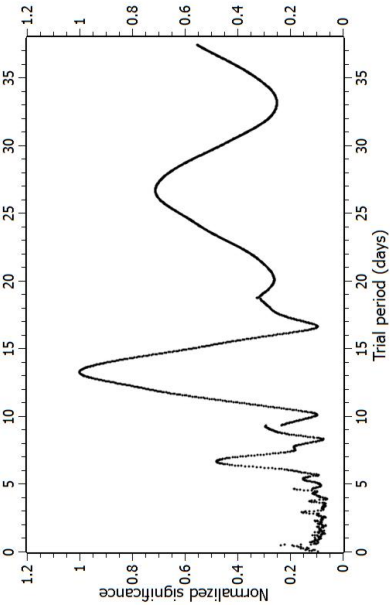
(α') Περιγραφή & ανιχνευμένες περιόδοι



(γ') Φάσμα FFT

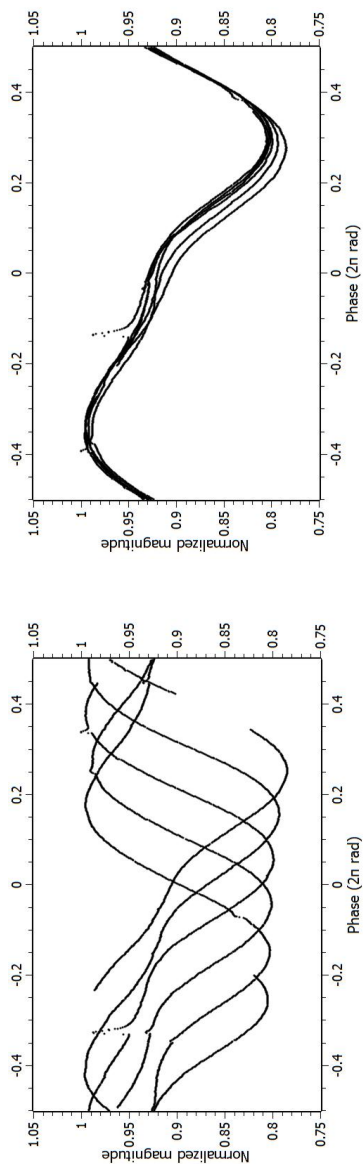


(β') Καμπύλη φωτός

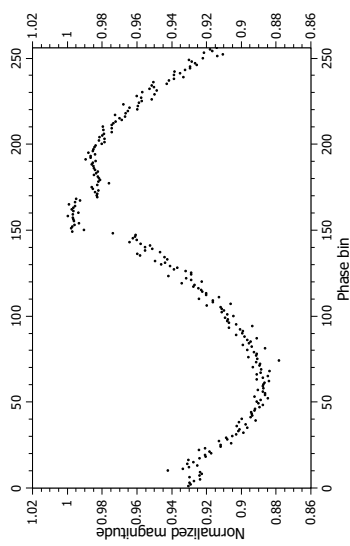


(δ') Περιοδογράμια FFA

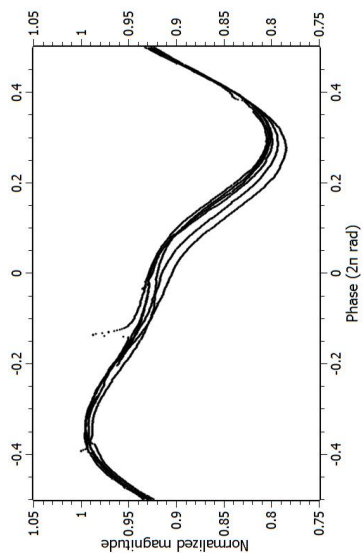
Σχήμα 3.16: Αποτελέσματα (α) για KID 12406908 — GSC 03555–00876



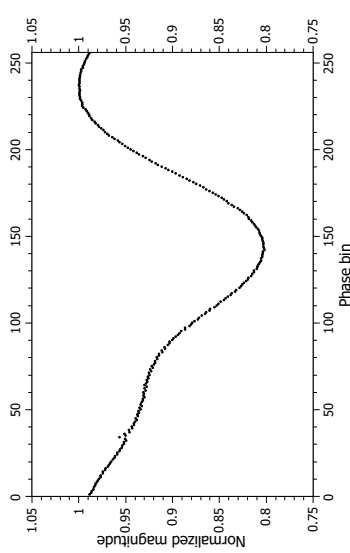
(α') Αναδίπλωση στην περίοδο του FFT



(γ') Re-binned αναδίπλωση στην περίοδο του FFT



(β') Αναδίπλωση στην περίοδο του FFA



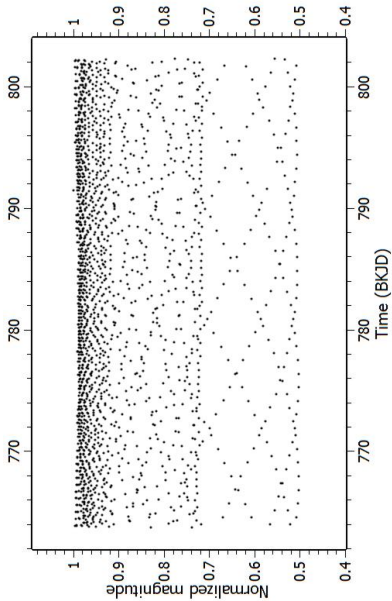
(δ') Re-binned αναδίπλωση στην περίοδο του FFA

Σχήμα 3.17: Αποτελέσματα (b) για KID 12406908 — GSC 03555–00876

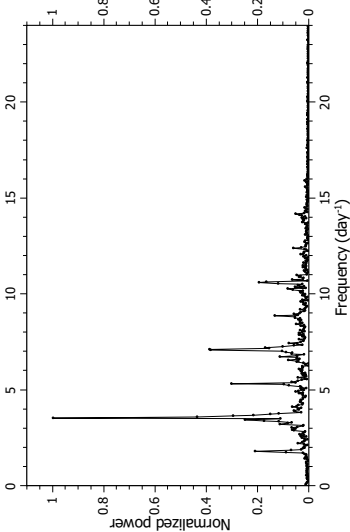
Kepler ID	4660997
Offset χρόνου	2454833 BJD
Τρίμηνο	8 (long)
Τύπος (ASAS)	Διπλός εκλειπτικός
Περίοδος (ASAS)	0.562560
Μετρήσεις	1862

Περίοδος FFT	0.283 346(1889)
Περίοδος FFA	0.280 887(82)

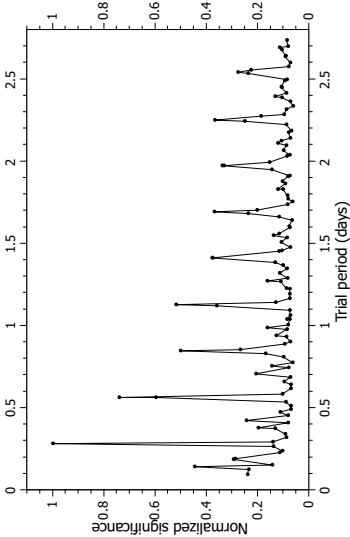
(α') Περιγραφή & ανιχνευμένες περιόδοι



(β') Καμπύλη φωτός

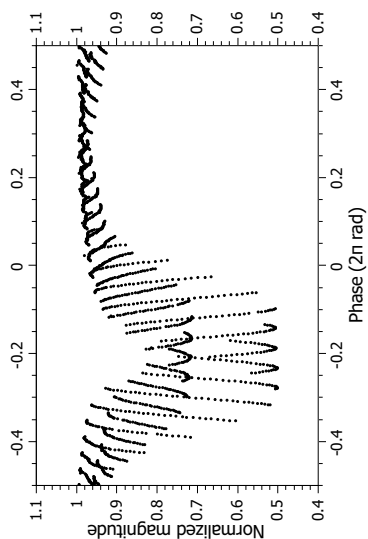


(γ') Φάσμα FFT

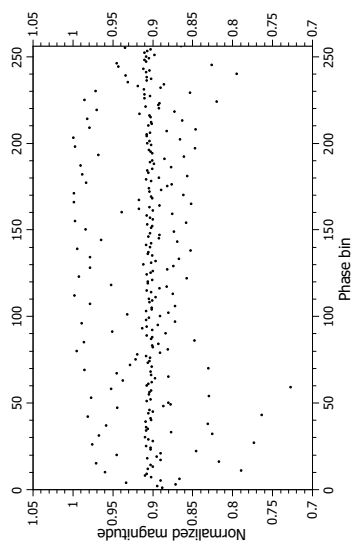


(δ') Περιοδογράμμο FFA

Σχήμα 3.18: Αποτελέσματα (α) για KID 4660997 — V1130 Cyg

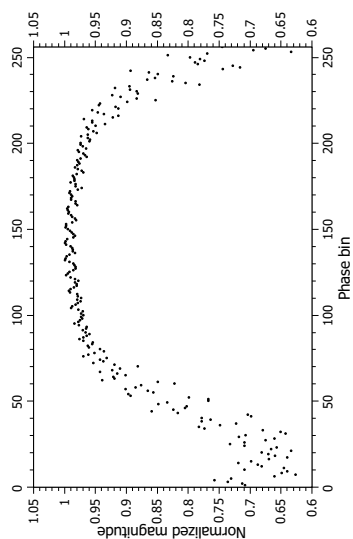


(α') Αναδίπλωση στην περίοδο του FFT



(γ') Re-binned αναδίπλωση στην περίοδο του FFT

(β') Αναδίπλωση στην περίοδο του FFA



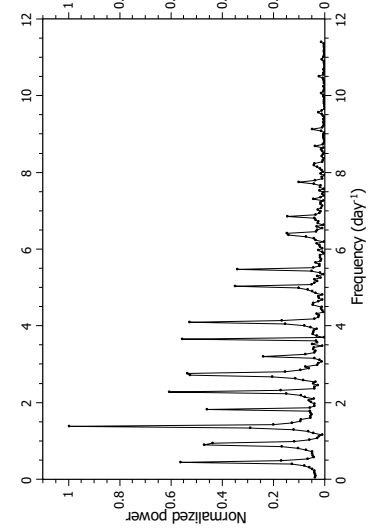
(δ') Re-binned αναδίπλωση στην περίοδο του FFA

Σχήμα 3.19: Αποτελέσματα (b) για KID 4660997 — V1130 Cyg

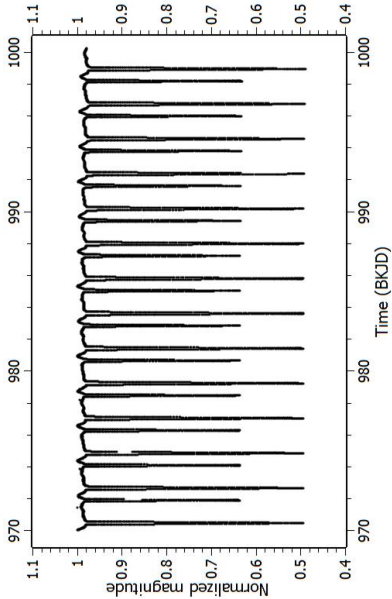
Kepler ID	4544587
Offset χρόνου	2454833 BJD
Τρίμηνο	10 (short)
Τύπος (ASAS)	Διπλός εκλειπτικός
Περίοδος (ASAS)	2.1893
Μετρήσεις	44062

Περίοδος FFT	0.724 834(11691)
Περίοδος FFA	0.730 418(11)

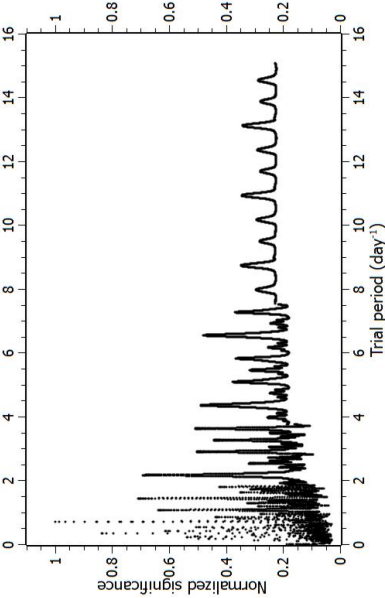
(α') Περιγραφή & ανιχνευμένες περίοδοι



(γ') Φάσμα FFT

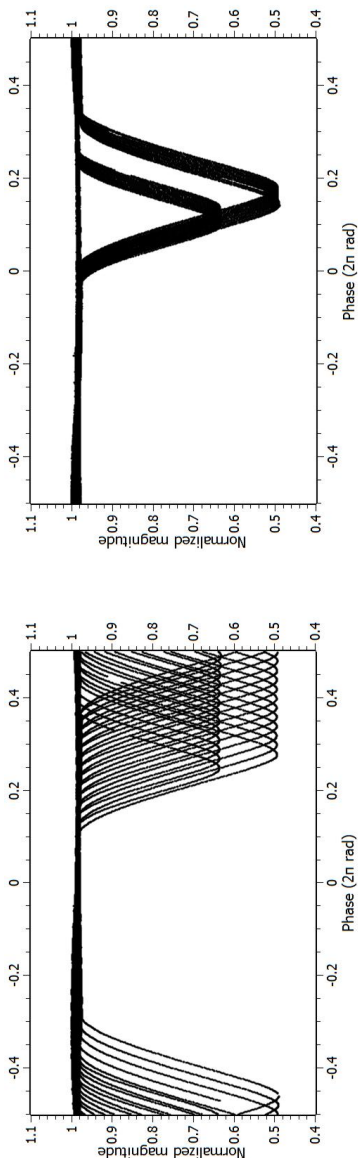


(β') Καμπύλη φωτός



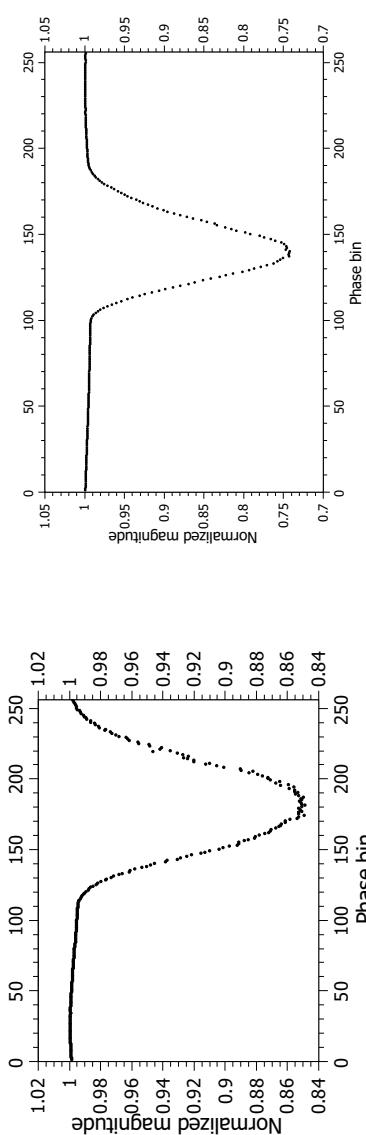
(δ') Περιοδογράμια FFA

Σχήμα 3.20: Αποτελέσματα (α) για KID 4544587 —TYC 3124–1348–1



(β') Αναδίπλωση στην περίοδο του FFA

(α') Αναδίπλωση στην περίοδο του FFT



(δ') Re-binned αναδίπλωση στην περίοδο του FFA

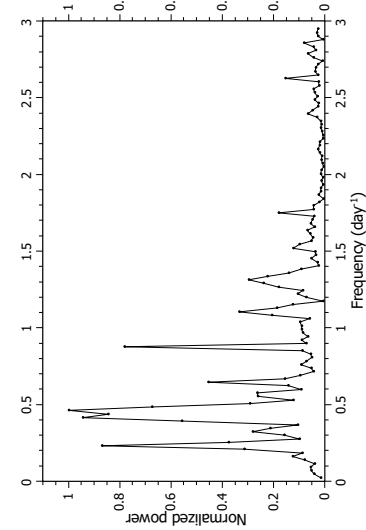
(γ') Re-binned αναδίπλωση στην περίοδο του FFT

Σχήμα 3.21: Αποτελέσματα (b) για KID 4544587 — TYC 3124-1348-1

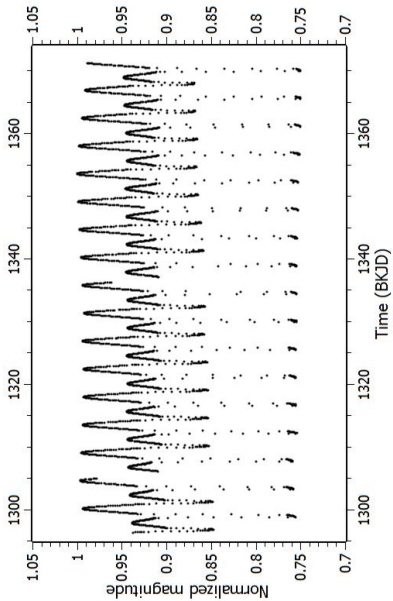
Kepler ID	8868650
Offset χρόνου	2454833 BJD
Τρίμηνο	14 (long)
Τύπος (ASAS)	Διπλός εκλειπτικός
Περίοδος (ASAS)	4.4472
Μετρήσεις	3535

Περίοδος FFT	2.172 630(54316)
Περίοδος FFA	2.222 982(342)

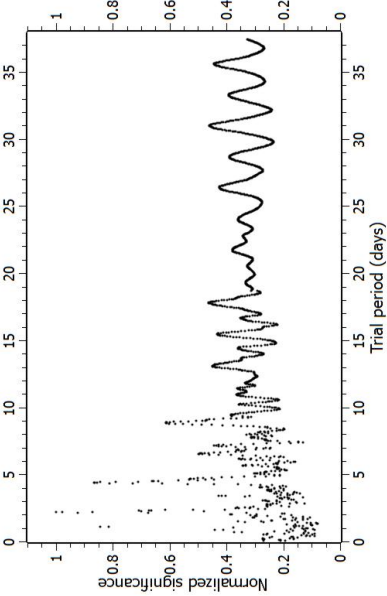
(α') Περιγραφή & ανιχνευμένες περίοδοι



(γ') Φάσμα FFT

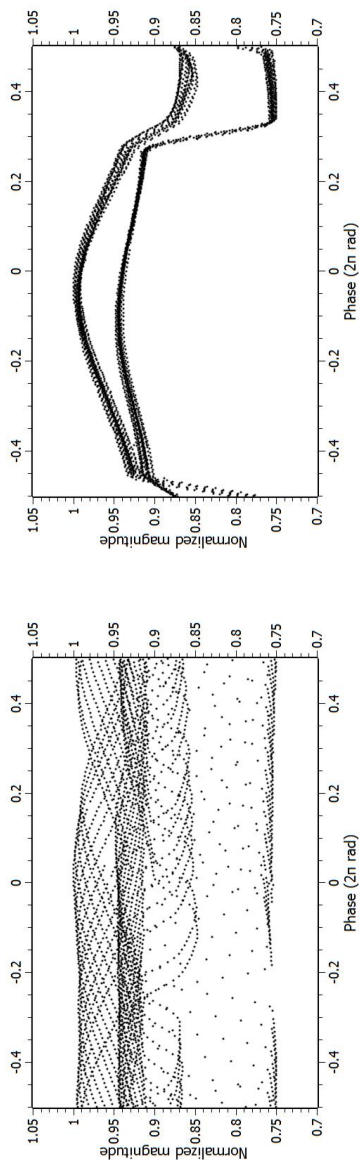


(β') Καμπύλη φωτός

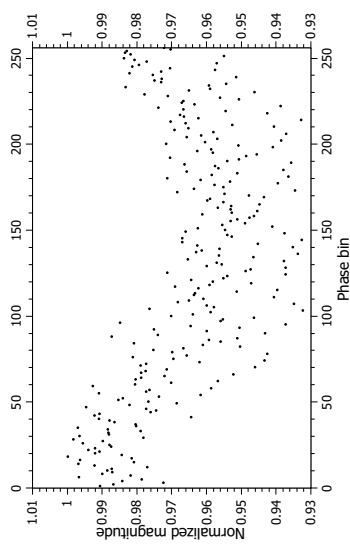


(δ') Περιοδογράμια FFA

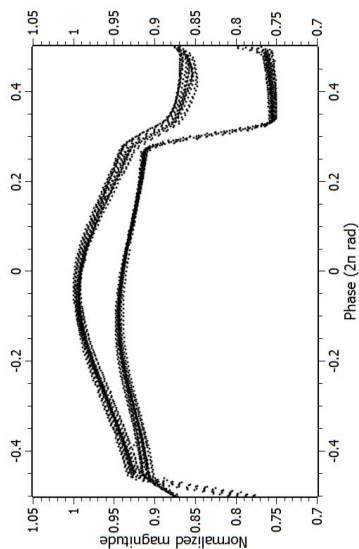
Σχήμα 3.22: Αποτελέσματα (α) για KID 8868650 — TYC 3541–2114–1



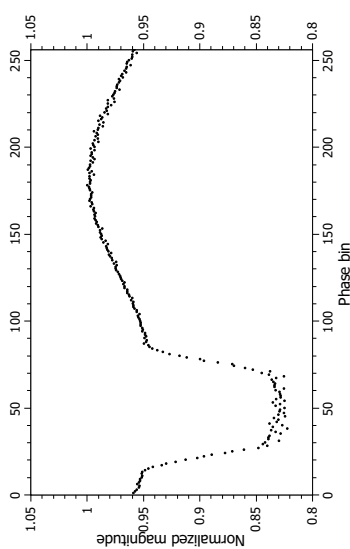
(α') Αναδίπλωση στην περίοδο του FFT



(γ') Re-binned αναδίπλωση στην περίοδο του FFT



(β') Αναδίπλωση στην περίοδο του FFA

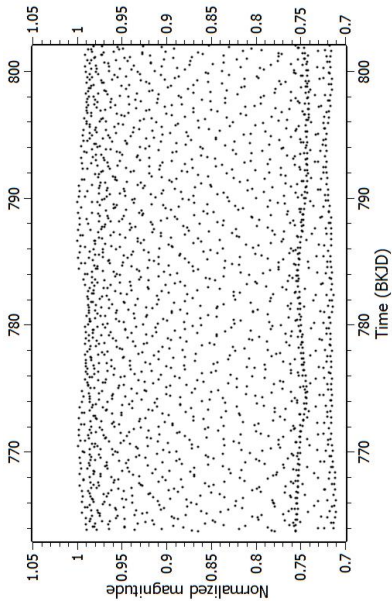


(δ') Re-binned αναδίπλωση στην περίοδο του FFA

Σχήμα 3.23: Αποτελέσματα (b) για KID 8868650 — TYC 3541-2114-1

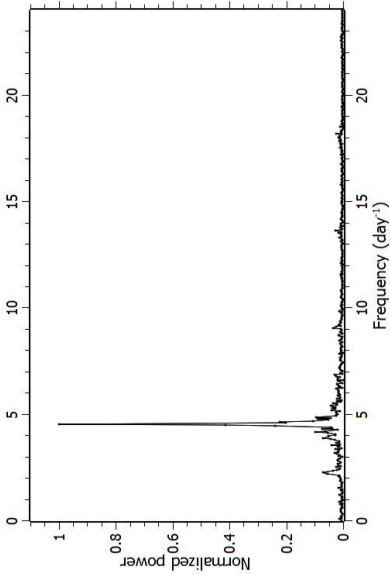
Kepler ID	10618251
Offset χρόνου	2454833 BJD
Τρίμηνο	8 (long)
Τύπος (ASAS)	Διπλός εκλειπτικός
Περίοδος (ASAS)	0.437417
Μετρήσεις	1863

Περίοδος FFT	0.219 075 (1129)
Περίοδος FFA	0.218 464 (82)

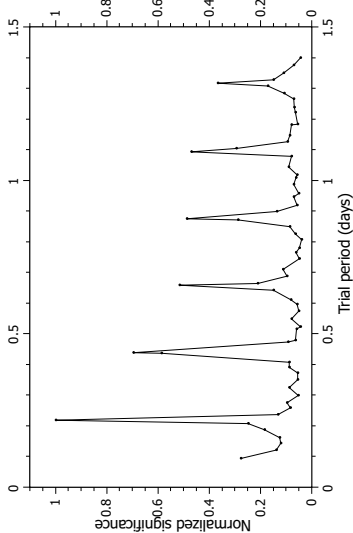


(α') Περιγραφή & ανιχνευμένες περίοδοι

(β') Καμπύλη φωτός

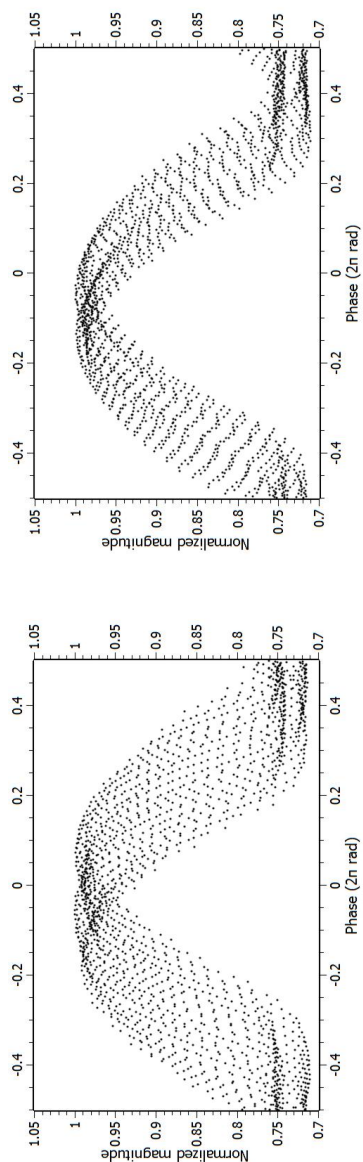


(γ') Φάσμα FFT

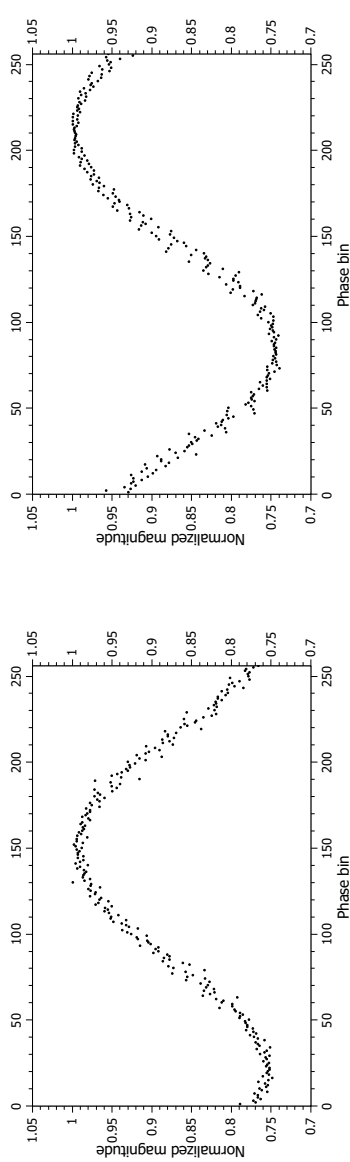


(δ') Περιοδόγραμμα FFA

Σχήμα 3.24: Αποτελέσματα (a) για KID 10618251 — ASAS 194725+4104.0



(β') Αναδίπλωση στην περίοδο του FFA

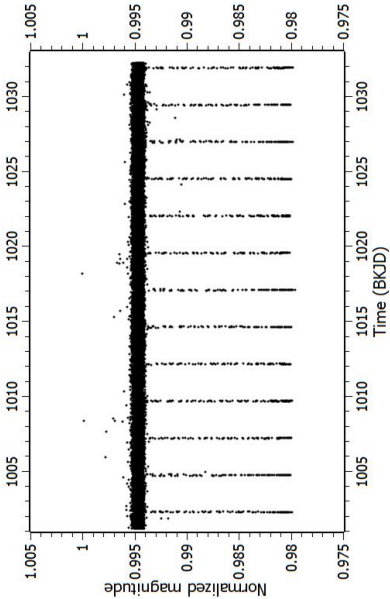


(δ') Re-binned αναδίπλωση στην περίοδο του FFA

(γ') Re-binned αναδίπλωση στην περίοδο του FFT

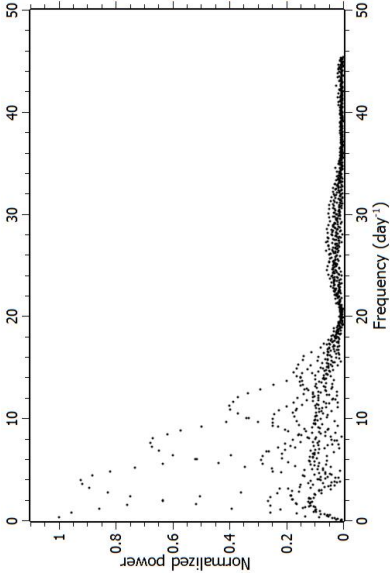
Σχήμα 3.25: Αποτελέσματα (b) για KID 10618251 — ASAS 194725+4104.0

Kepler ID	11446443
Offset χρόνου	2454833 BJD
Τρίμηνο	11 (short)
Τύπος (Kepler)	Host of exoplanet
Περίοδος (Kepler)	2.47063
Μετρήσεις	45183
Περίοδος FFT	2.510 362 (139465)
Περίοδος FFA	0.003 477 (1)

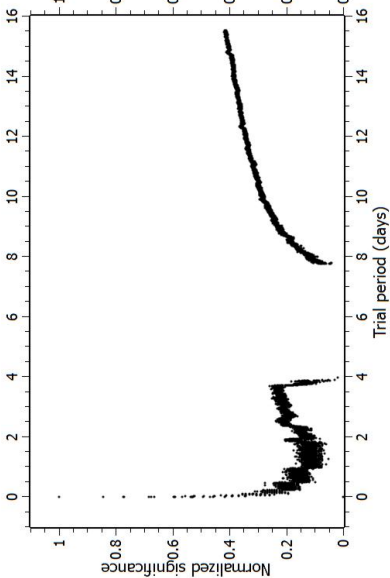


(α') Περιγραφή & ανιχνευμένες περιόδοι

(β') Καμπύλη φωτός

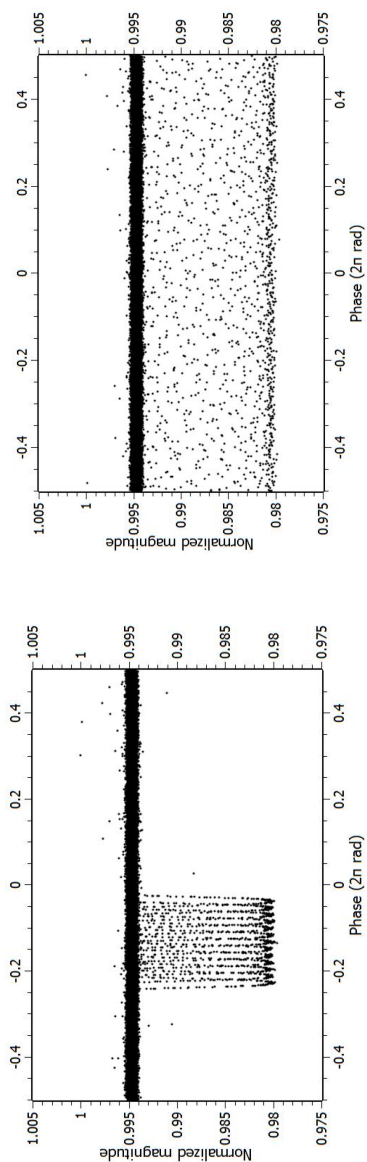


(γ') Φάσμα FFT



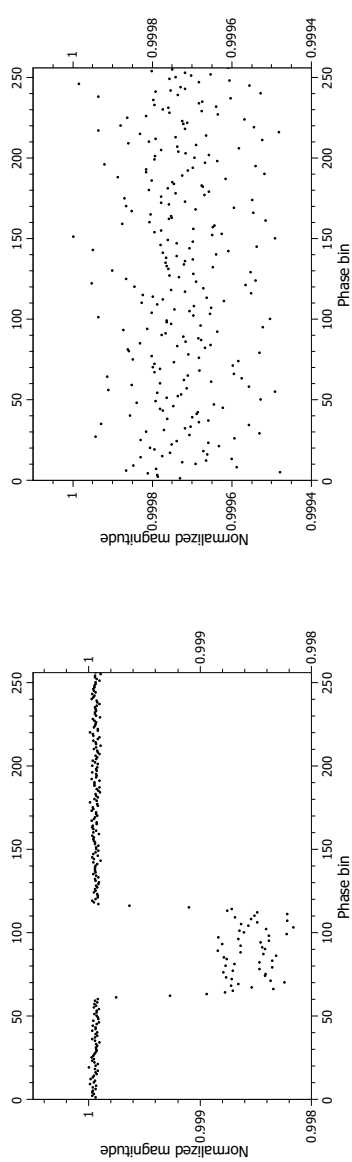
(δ') Περιοδόγραμμα FFA

Σχήμα 3.26: Αποτελέσματα (a) για KID 11446443 — KOI-1 — TYC 3549-2811-1



(α') Αναδίπλωση στην περίοδο του FFT

(β') Αναδίπλωση στην περίοδο του FFA



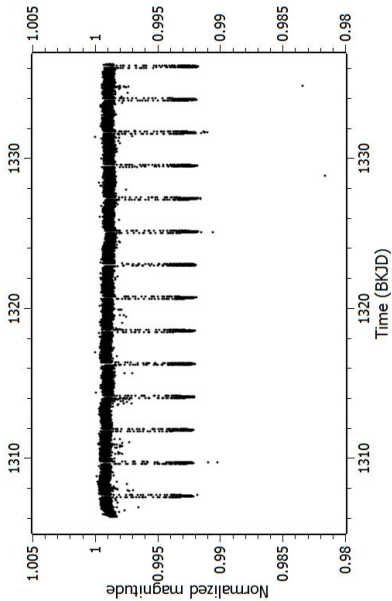
(γ') Re-binned αναδίπλωση στην περίοδο του FFT

(δ') Re-binned αναδίπλωση στην περίοδο του FFA

Σχήμα 3.27: Αποτελέσματα (b) για KID 1144643 — KOI-1 — TYC 3549-2811-1

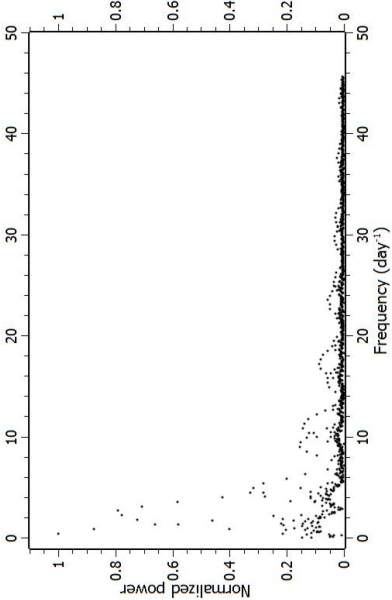
Kepler ID	10666592
Offset χρόνου	2454833 BJD
Τρίμηνο	14.2 (short)
Τύπος (Kepler)	Host of exoplanet
Περίοδος (Kepler)	2.204736
Μετρήσεις	44149

Περίοδος FFT	2.244 355(112218)
Περίοδος FFA	0.003 497(1)

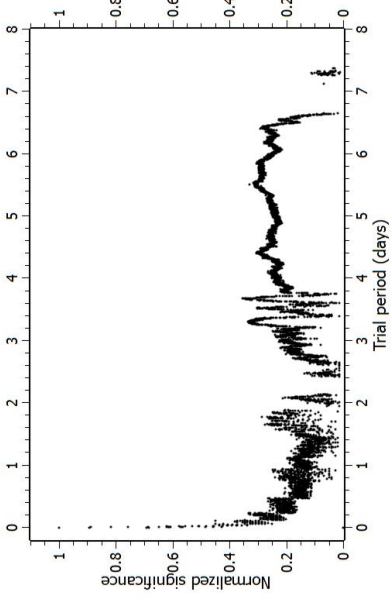


(α') Περιγραφή & ανιχνευμένες περιόδοι

(β') Καμπύλη φωτός

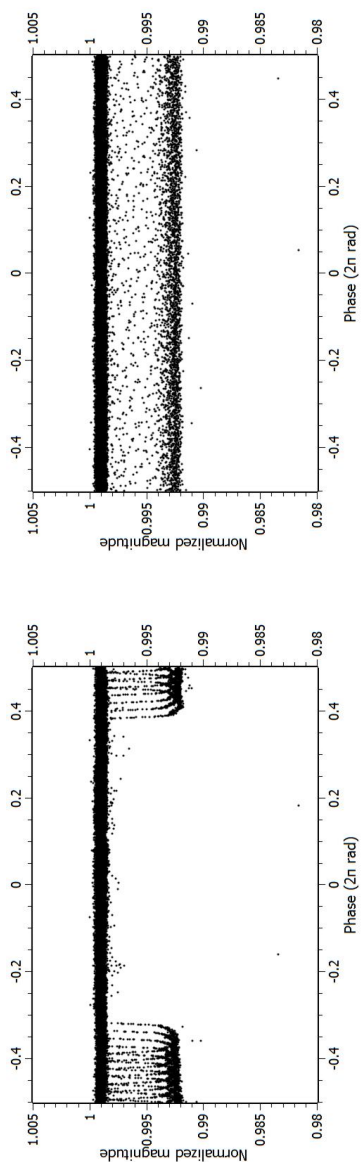


(γ') Φάσμα FFT



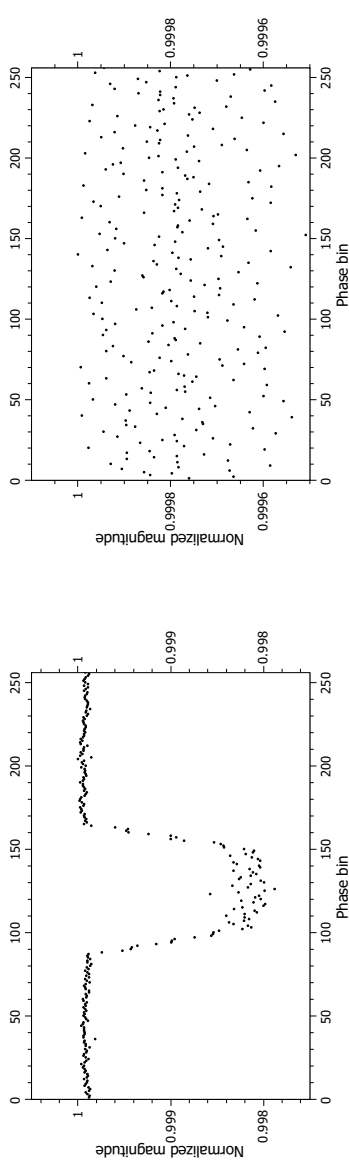
(δ') Περιοδόγραμμα FFA

Σχήμα 3.28: Αποτελέσματα (α) για KID 10666592 — KOI-2 — BD+47 2846



(α') Αναδίπλωση στην περίοδο του FFT

(β') Αναδίπλωση στην περίοδο του FFA



(γ') Re-binned αναδίπλωση στην περίοδο του FFT

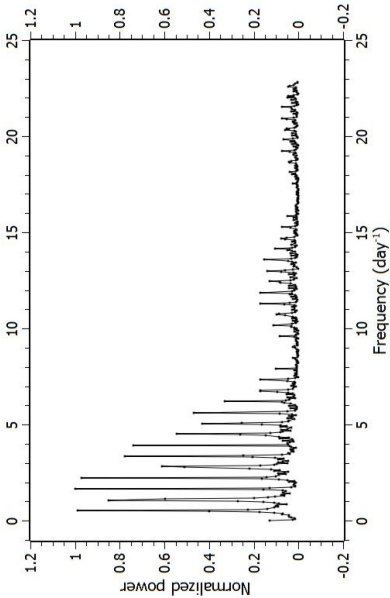
(δ') Re-binned αναδίπλωση στην περίοδο του FFA

Σχήμα 3.29: Αποτελέσματα (b) για KID 10666592 — KOI-2 — BD+47 2846

Kepler ID	9941662
Offset χρόνου	2454833 BJD
Τρίμηνο	14.3 (short)
Τύπος (Kepler)	Host of exoplanet
Περίοδος (Kepler)	1.7637
Μετρήσεις	49958

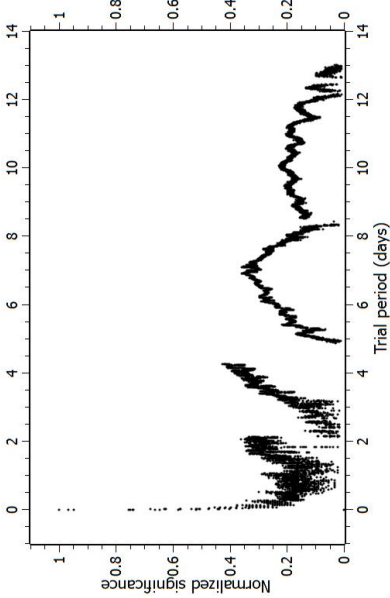
Περίοδος FFT	0.589 626(7758)
Περίοδος FFA	0.002 735(1)

(α') Περιγραφή & ανιχνευμένες περίοδοι



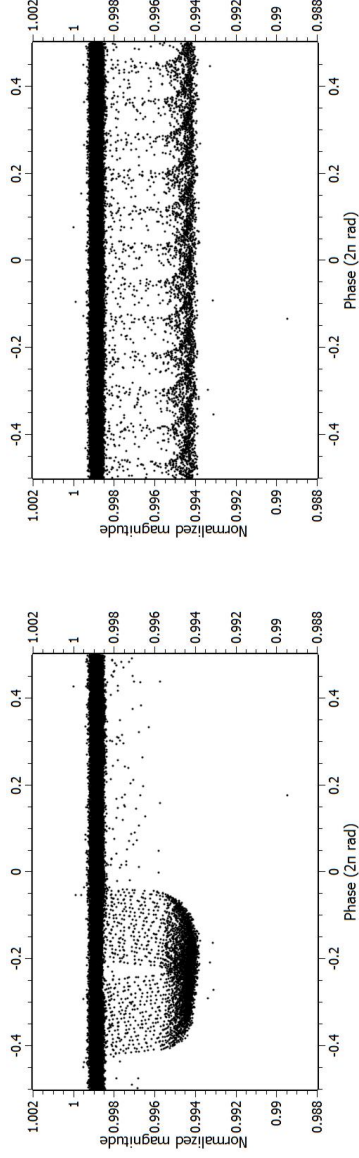
(γ') Φάσμα FFT

(β') Καμπύλη φωτός



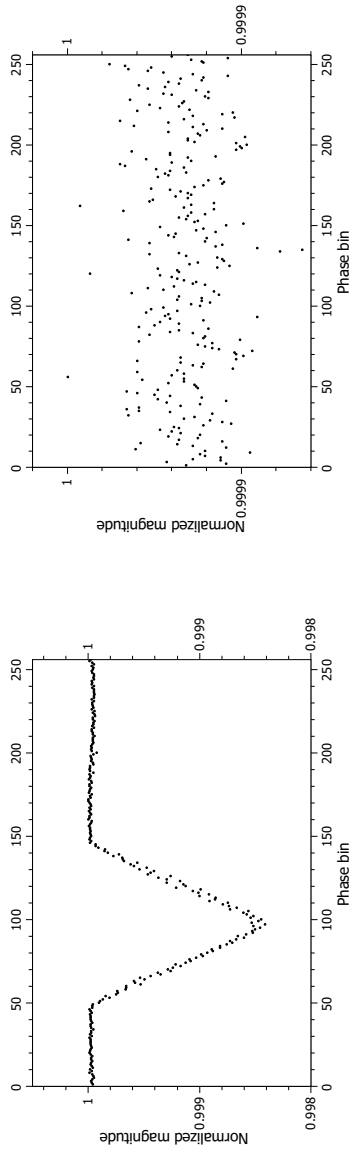
(δ') Περιοδόγραμμα FFA

Σχήμα 3.30: Αποτελέσματα (a) για KID 9941662 — KOI-13 —BD+46 2629



(α') Αναδίπλωση στην περίοδο του FFT

(β') Αναδίπλωση στην περίοδο του FFA



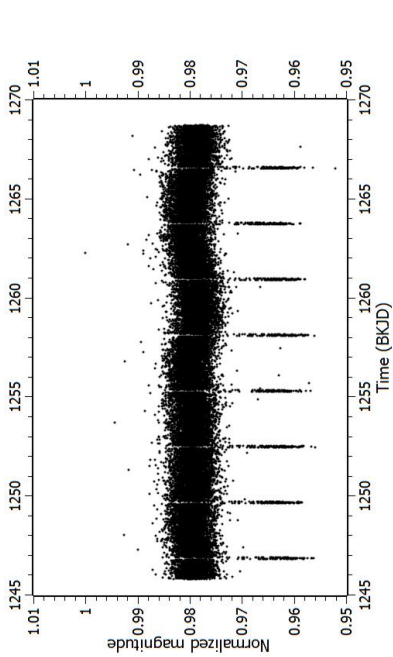
(γ') Re-binned αναδίπλωση στην περίοδο του FFT

(δ') Re-binned αναδίπλωση στην περίοδο του FFA

Σχήμα 3.31: Αποτελέσματα (b) για KID 9941662 — KOI-13 — BD+46 2629

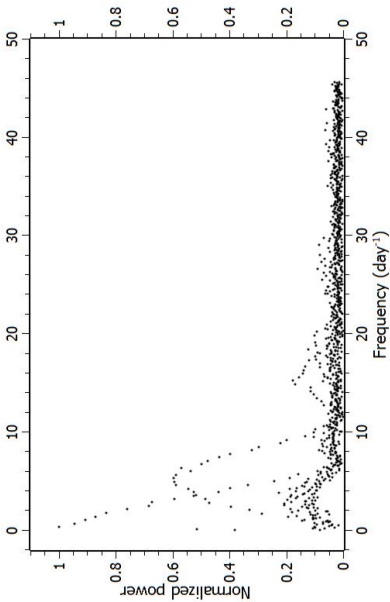
Kepler ID	11414511
Offset χρόνου	2454833 BJD
Τρίμηνο	13.3 (short)
Τύπος (Kepler)	Host of exoplanet
Περίοδος (Kepler)	2.816508
Μετρήσεις	33464

Περίοδος FFT	2.811 015(175689)
Περίοδος FFA	0.002 961(1)

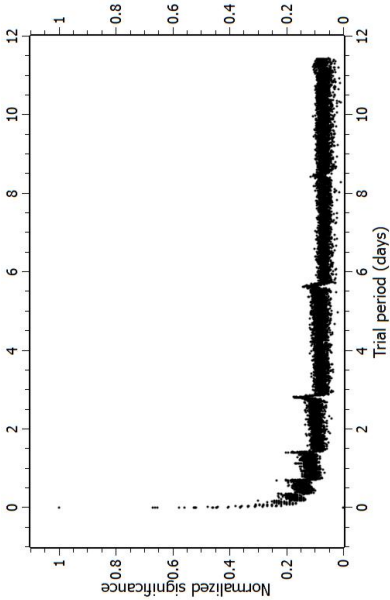


(β') Καμπύλη φωτός

(α') Περιγραφή & ανιχνευμένες περιόδοι

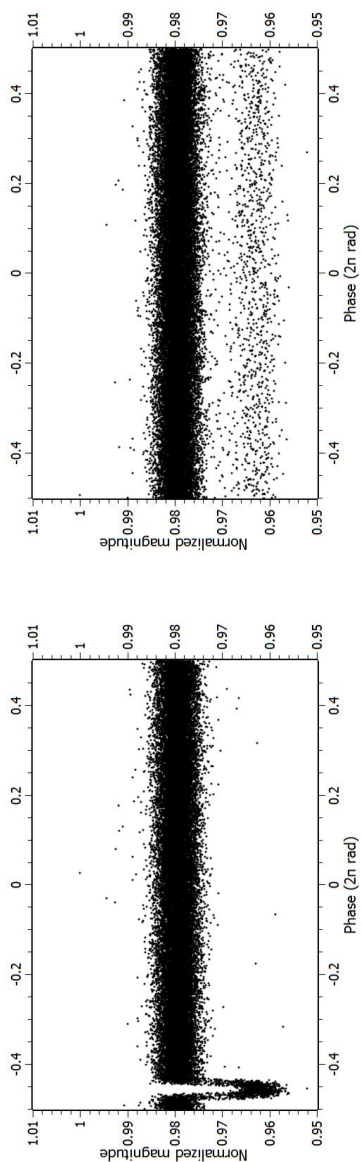


(γ') Φάσμα FFT



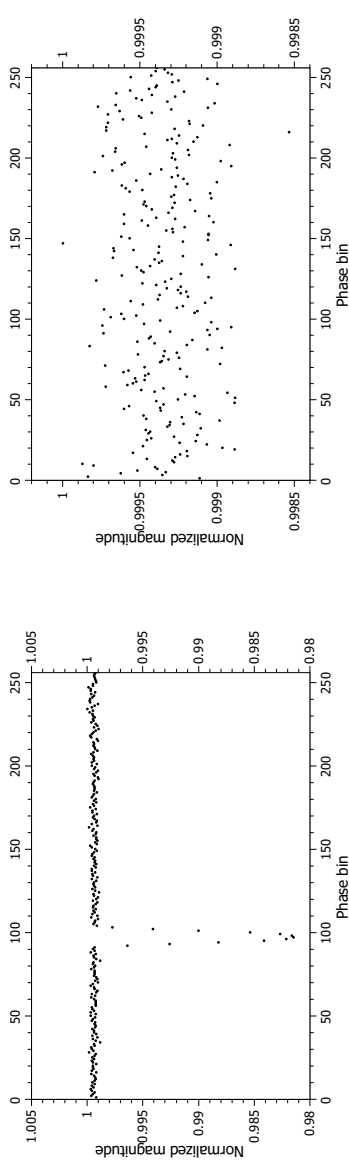
(δ') Περιοδόγραμμα FFA

Σχήμα 3.32: Αποτελέσματα (α) για KID 11414511 — KOI-767 — 2MASS J19480371+4913310



(α') Αναδίπλωση στην περίοδο του FFT

(β') Αναδίπλωση στην περίοδο του FFA



(γ') Re-binned αναδίπλωση στην περίοδο του FFT

(δ') Re-binned αναδίπλωση στην περίοδο του FFA

Σχήμα 3.33: Αποτελέσματα (b) για KID 11414511 — KOI-767 — 2MASS J19480371+4913310

4

Συζήτηση

4.1 Τεστ υπολογιστικής πολυπλοκότητας

FFT

Ο υπολογιστής στον οποίο έτρεξε το πρόγραμμα ήταν αργός σε σχέση με σύγχρονα εργαστηριακά τεματικά (Pentium® Dual-Core CPU στα $2.2GHz$, με Microsoft Windows® 7 64-bit). Παρά το γεγονός αυτό, ο FFT έδειξε το μεγάλο πλεονέκτημά του: την ταχύτητα.

Ακόμα και με είσοδο 16.8 εκατομμυρίων στοιχείων (2^{24}), ο υπολογισμός του διακριτού μετασχηματισμού Fourier χρειάστηκε λίγο λιγότερο από ένα λεπτό. Με άλλα λόγια, θα χρειαζόντουσαν 60 μόνο δευτερόλεπτα για να την ανάλυση μίας συλλογής μετρήσεων που λήφθηκαν με διαφορά ενός δευτερολέπτου για περίπου 6.5 μήνες.

Η προσαρμογή των χρονομετρήσεων σε μία καμπύλη τύπου $a_1 + a_2 N \log_2 N$

επέστρεψε συντελεστή αποφασιστικότητας R^2 ίσο με 0.99287, επιβεβαιώνοντας την πολυπλοκότητα του FFT και την καλή λειτουργία της δικιάς μας πραγματοποίησης του αλγορίθμου.

Ωστόσο, μία προσεκτική εξέταση του συντελεστή (σταθεράς) του πιο σημαντικού όρου ($N \log_2 N$) αποκαλύπτει μία μη αναμενόμενη συμπεριφορά: η τιμή του αυξάνει σε σκάλες για $N = 2^{16}$ και $N = 2^{24}$. Οι αριθμοί υποδεικνύουν συνολική χρήση μνήμη που ξεπρανά διάφορα επίπεδα CPU cache¹ (L2, L3, RAM), οδηγώντας στη χρήση ολοένα και αργότερων μνημών του συστήματος.

FFA

Με πολυπλοκότητα $O(n^2)$, ο FFA είναι εξαιρετικά αργός. Για μεγέθη εισόδου χιλιάδων στοιχείων — μία λογική τιμή για αστρονομικά δεδομένα — ο FFT ολοκλήρωσε την ανάλυση εκατοντάδες φορές πιο γρήγορα.

Επίσης, η προσαρμογή των χρονομετρήσεων στο πολυώνυμο $a_0 + a_1 N + a_2 N^2$ επέστρεψε συντελεστή αποφασιστικότητας R^2 ίσο με 0.99985 αποδεικνύοντας τις θεωρητικές μας εκτιμήσεις (βλ. Παράρτημα Α'.2).

Η επίδραση των επιπέδων cache μνήμης δεν έγινε εμφανής στην περίπτωση του FFA. Αυτό οφείλεται στη μη αναδρομική πραγματοποίηση του αλγορίθμου, γεγονός που μειώνει την απαιτούμενη μνήμη κατά έναν παράγοντα 2. Επίσης, καθώς δε φτάσαμε ποτέ σε μεγάλες εισόδους χρησιμοποιήθηκε η L2 cache σε όλες τις περιπτώσεις.

4.2 Τεστ σε κατασκευασμένες καμπύλες φωτός

Ημιτονοειδή σήματα

Στην περίπτωση των ημιτονοειδών καμπύλων φωτός (Σχήμα 3.5(α')), ο FFT ανίχνευσε επιτυχώς όλα τα σήματα με ακρίβεια μικρότερη του 10%. Υπήρξε μόνο μία εξαίρεση στους χιλιάδες συνδυασμούς περιόδων και επιπέδων θορύβου, και

¹ Η μνήμη Cache L2 ήταν 1MB. Για τέσσερις πίνακες (είσοδος, έξοδος, δύο buffer για την αναδρομική κλήση) των 2^{16} μετρήσεων, με 4-byte έκαστη (λέξη υπολογιστή), χρειαζόμασταν ακριβώς 1MB ελεύθερης cache

αντιστοιχούσε σε μία καμπύλη με υψηλό θόρυβο (τυπική απόκλιση 3 φορές το πλάτος του σήματος). Το αποτέλεσμα αυτό δεν έρχεται σε αντίθεση με αυτό που περιμέναμε: ο FFT είναι κατάλληλος για ημιτονοειδή σήματα.

Αντιθέτως, ο FFA απέτυχε σε πολλές περιπτώσεις. Στο Σχήμα 3.5(β') μπορούμε να δούμε μία καμπύλη στο επίπεδο περιόδου – θορύβου που ξεχωρίζει περιοχές όπου κυριαρχούν επιτυχίες και αποτυχίες ανίχνευσης. Η καλή λειτουργία του FFA για χαμηλό θόρυβο είναι τετριμμένη. Η τάση να συμπεριφέρεται καλά για μικρότερες περιόδους επιβεβαιώνει ότι ο λόγος σήματος – θορύβου αυξάνεται με τον αριθμό των αναδιπλώσεων.

Εξετάζοντας τα σφάλματα των εκτιμήσεων περιόδου για τις επιτυχίες τους κάθε αλγορίθμου (Σχήμα 3.6) παρατηρήσαμε μία βασική διαφορά μεταξύ τους. Ο FFT εμφανίζει μεγαλύτερα σφάλματα για μικρές συχνότητες εξαιτίας της φασματικής διάχυσης και ο θόρυβος παίζει ελαχιστο ρόλο στις επιδόσεις του. Μία εξήγηση για το γεγονός αυτό είναι ότι το φάσμα του Γκαουσιανού θορύβου (Γκαουσιανός θόρυβος) είναι επίπεδο, μοιράζοντας την ισχύ σε όλες τις συχνότητες, καθιστώντας τα όποια μέγιστα ευδιάκριτα.

Εν αντιθέσει, ο FFA είχε διαφορετική συμπεριφορά (Σχήμα 3.6(β')). Η διαδικασία της αναδίπλωσης δεν εξουδετερώνει μεγάλο μέρος του θορύβου κάνοντας τον αλγόριθμο ευαίσθητο στην παρουσία του. Στον άλλον άξονα, η περίοδος δε φαίνεται να επιδρά σημαντικά στην επιτυχία του FFA.

Ορθογώνια σήματα

Στην περίπτωση σημάτων που έχουν μικρή ομοιότητα με τα ημιτονοειδή, περιμέναμε να δούμε πλεονέκτημα του αλγορίθμου FFA απέναντι στον FFT. Στα Σχήματα 3.7 to 3.9 παρατηρούμε ότι για ολοένα και μικρότερους κύκλους εργασίας δυσκολεύει η εύρεση της περιόδου, κάτι αναμενόμενο. Αυτό που δεν περιμέναμε ήταν ότι ο FFA ποτέ δεν είχε μεγαλύτερη επιτυχία από τον FFT. Ακόμα και για κύκλο εργασίας 5%, ο Fourier ήταν επιτυχής σε διπλάσιες στο πλήθος περιπτώσεις.

Προσπαθώντας να εξηγήσουμε την αντίφαση αυτή, πραγματοποιήσαμε διάφορα τεστ με χρήση μαθηματικών πακέτων ώστε να ελεγχθεί η έξοδος των αλγορίθμων και να εντοπιστούν πιθανά προβλήματα. Τα τεστ έδειξαν ότι οι αλγόριθμοι

μοι ήταν εξαιρετικά ακριβείς.

Η αποτυχία του FFA μπορεί να ερμηνευθεί από την κακή επιλογή μέτρου σημαντικότητας. Όπως αναφέρεται στην § 2.1.2, από αυτό το μέτρο διαλέγουμε τη βασικότερη περιοδικότητα του σήματος. Εξαρτάται από την ισχύ της εκάστοτε αναδίπλωσης. Κανένας υπολογισμός για την ισχύ του θορύβου δεν πραγματοποιείται ενώ το μέτρο λαμβάνει υπόψη μόνο μία εκτίμηση της βελτίωσης του λόγου σήματος – θορύβου (λόγω του κεντρικού οριακού θεωρήματος και της φύσης του Γκαουσιανού θορύβου).

Έτσι, για μικρούς κύκλους εργασίας ή/και για υψηλό θόρυβο, η ολική ισχύς προσεγγίζει την ισχύ του θορύβου, οδηγώντας σε αποτυχίες ανίχνευσης. Μία πιο κατάλληλη μέθοδος εκτίμησης του λόγου σήματος – θορύβου θα έπρεπε να χρησιμοποιηθεί (όπως για παράδειγμα windowing functions ή σύγκριση με φιλτραρισμένες εκδοχές της αναδίπλωσης).

4.3 Ανάλυση δεδομένων του Kepler

Γενική εικόνα

Στον Πίνακα 3.2 μπορούμε να δούμε ότι και οι δύο αλγόριθμοι κατάφεραν να εντοπίσουν την περίοδο όλων των Κηφείδων και τον διπλων εκλειπτικών με σφάλμα μικρότερο από 10% σε σχέση με τη δημοσιευμένη τιμή από την ομάδα του All Sky Automated Survey (Pigulski et al., 2009). Στην περίπτωση των διαβαίνοντων εξωηλιακών πλανητών, το προηγούμενως αναφερόμενο πρόβλημα του μέτρου σημαντικότητας του FFA κατέστησε τις διαβάσεις ανίκανες να ανιχνευθούν, ενώ ο FFT εξετέλεσε το καθήκον του δίνοντας πολύ καλές εκτιμήσεις σε σχέση με τις δημοσιευμένες περιόδους των εν λόγω στόχων (O'Donovan et al., 2006; Welsh et al., 2010; Shporer et al., 2011; Borucki et al., 2011).

Εδώ θα πρέπει να σημειώσουμε ότι το τηλεσκόπιο Kepler παρείχε υψηλής ακρίβειας δεδομένα με πολύ μεγάλο λόγο σήματος θορύβου επιτρέποντας στον FFT να εντοπίσει τις διαβάσεις.

Αναδίπλωση σε εκτιμήσεις της περιόδου

Τα σφάλματα των εκτιμήσεων της περιόδου σε σχέση με τις δημοσιευμένες — και πιο έμπιστες — τιμές του ASAS (Pigulski et al., 2009), αντιστοιχούν σε αποκλίσεις συγκρινόμενες σε μέγεθος με το ρυθμό δειγματοληψίας. Η αναδίπλωση χωρίς re-binning αποκαλύπτει το γεγονός αυτό με την εμφάνιση πολλαπλών γειτονικών προφίλ. Αυτό ισχύει για όλες τις περιπτώσεις, εκτός από πέντε: (i) αναδίπλωση με τον FFT για τον KID 4660997 (Σχήμα 3.19(α')), όπου η απόκλιση είναι μικρότερη κατά 5 περίπου φορές από την περίοδο δειγματοληψίας και πάνω από 60 αναδιπλώσεις έχουν εφαρμοστεί, και στην περίπτωση (ii) αναδίπλωση για την περίοδο του FFA για όλους τους εξωπλανήτες όπου η ανίχνευση απέτυχε.

Εξαιτίας αυτού του γεγονότος, η re-binned αναδίπλωση σε περίοδο κοντά στην πραγματική, οδηγεί σε απαλές και διευρυμένες κορυφές στην καμπύλη φωτός. Ξεκάθαρη απεικόνιση αυτού έχουμε στην περίπτωση των Σχήμα 3.11(γ') and Σχήμα 3.11(δ'). Παράδειγμα της διεύρυνσης έχουμε για όλους τους εξωπλανήτες με τον FFT.

Ωστόσο, τα σφάλματα του FFA ήταν σημαντικά μικρότερα του FFT, και οι re-binned αναδιπλώσεις για τους Κηφειδες μεταβλητούς αστέρες και τα διπλά εκλειπτικά συστήματα μπορούν να θεωρηθούν αρκετά ακριβείς.

Μία καλή περίπτωση διάβασης

Η περίοδος περιφοράς ενός εξωπλανήτη περί του αστέρα KID 11414511 ανιχνεύθηκε με πολύ μικρό σφάλμα από τον FFT και η σχετικά μεγάλη περίοδος (συγκρινόμενη με το ρυθμό δειγματοληψίας) οδήγησε σε ένα πολύ ακριβές προφίλ: δημοσιευμένες τιμές για την περίοδο και τη διάρκεια της διάβασης (Boruckiet al., 2011) υποδεικνύουν κύκλο εργασίας 3.8%, ενώ μελετώντας το Σχήμα 3.33(δ') βρήκαμε κύκλο εργασίας 5.1% (διάρκεια: 13 bins φάσης από τα 256).

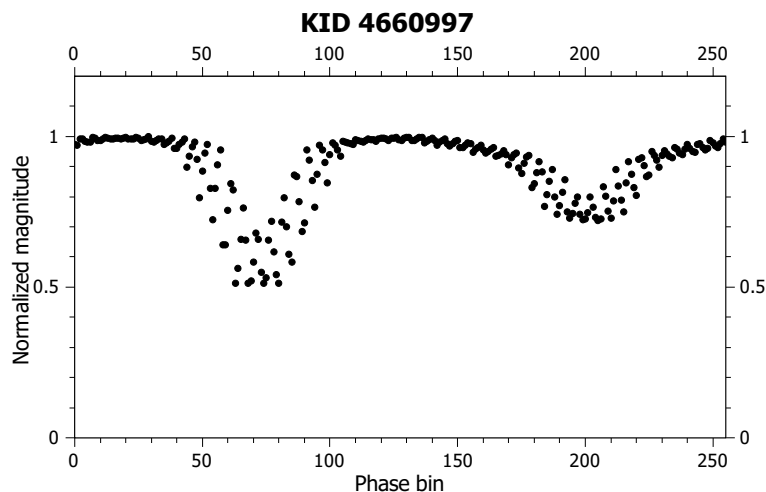
Ανίχνευση εσφαλμένη αρμονικής

Στην περίπτωση των διπλών εκλειπτικών, δύο παρόμοιας μορφής εκλείψεις εμφανίζονται σε κάθε περίοδο. Πολλές φορές, το διάστημα μεταξύ δύο ελάχιστων είναι ίσο με το μισό της περιόδου, κάτι που οδηγεί σε μέγιστο της ισχύος

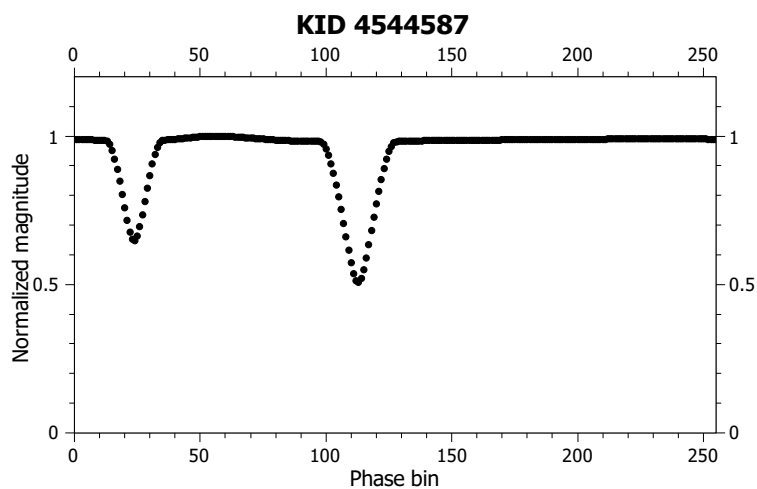
στην αντίστοιχη συχνότητα, στον DFT. Αρκετά έκκεντρες τροχιές με το μεγάλο ημιάξονα σχεδόν κάθετο στην ευθεία παρατήρησης μπορεί να προκαλέσουν μία διαφορετική κατάτμηση του χρόνου ανάμεσα σε πρωτεύουσες και δευτερεύουσες εκλείψεις. Για παράδειγμα για 2:1 κατάτμηση, έχουμε ανίχνευση περιόδου που αντιστοιχεί σε αρμονική περιόδου 3 φορές μικρότερης της πραγματικής.

Για να αντιμετωπίσουμε το πρόβλημα αυτό, δημιουργήσαμε τις μη re-binned αναδιπλώσεις όπου όλες οι παρατηρήσεις αντιστοιχούν στο χρόνο παρατήρησης modulo την περίοδο. Έτσι, σε περίπτωση εσφαλμένης αρμονικής μπορούσαμε να δούμε όλα τα προφίλ.

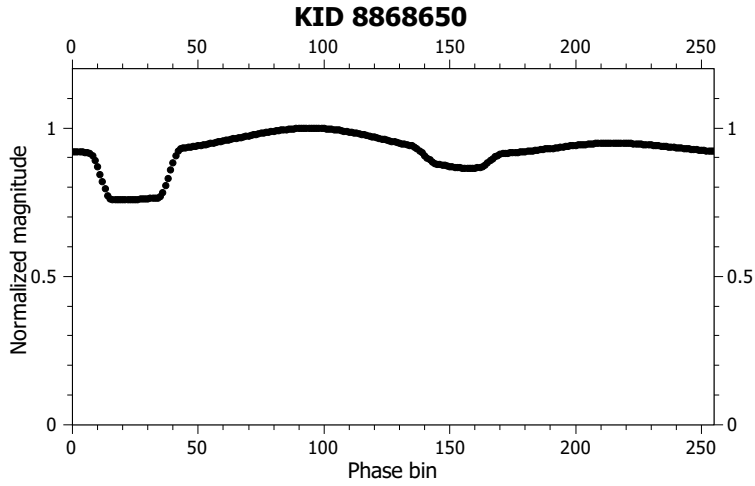
Οι εκτιμήσεις της περιόδου στον Πίνακα 3.2 είναι οι διορθωμένες. Ωστόσο, στους πίνακες και τα σχήματα έχουμε τις πραγματικές τιμές. Πραγματοποιήσαμε re-binned αναδιπλώσεις, Σχήματα 4.1 to 4.4 και για τους τέσσερις στόχους, στη διορθωμένη περίοδο που εντόπισε ο FFA (καθώς ήταν και πιο ακριβής) για να αποκαλύψουμε την πραγματική μορφή των καμπυλών φωτός των διπλών εκλείπτικών.



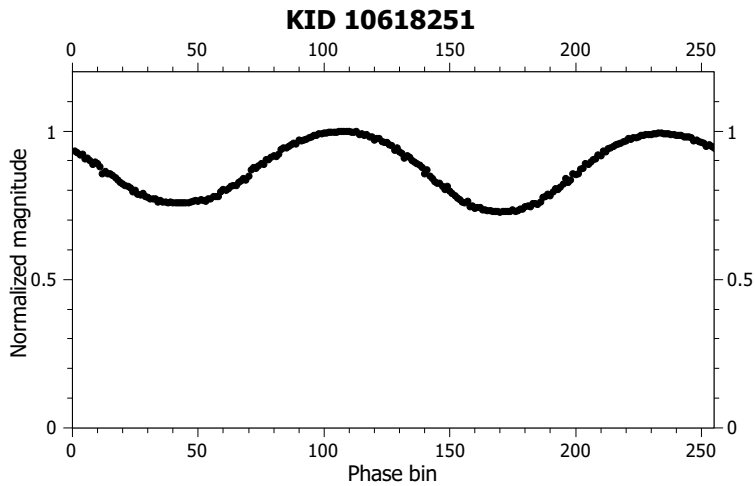
Σχήμα 4.1: Διορθωμένη καμπύλη φωτός για τον τύπου Algol-type διπλό εκλειπτικό KID 4660997. Η περίοδος της αναδίπλωσης διπλασιάστηκε για να απεικονιστούν και οι δύο εκλείψεις. Η πολύ μικρή περίοδος (13^h36^m) είναι συγκρίσιμη με την περίοδο δειγματοληψίας (30^m) οδηγώντας σε μεγάλες αποκλίσεις που εμφανίζονται ως διάχυση στο σχήμα.



Σχήμα 4.2: Διορθωμένη καμπύλη φωτός για τον τύπου Algol διπλό εκλειπτικό KID 4544587. Η περίοδος της αναδίπλωσης **τριπλασιάστηκε** για να απεικονιστούν και οι δύο εκλείψεις αυτού του συστήματος με μεγάλη εκκεντρότητα. Αποτελεί παράδειγμα του πόσο προσεκτικοί πρέπει να είμαστε ερμηνεύοντας τις αρμονικές μίας χρονοσειράς.



Σχήμα 4.3: Διορθωμένη καμπύλη φωτός για τον τύπου Algol διπλό εκλειπτικό KID 8868650. Η περίοδος της αναδίπλωσης διπλασιάστηκε για να απεικονιστούν και οι δύο εκλείψεις. Παρατηρούμε το τετραγωνισμένο προφίλ των ελάχιστων, πιθανότατα λόγω του μικρού μεγέθους του συνοδού αστέρα, ο οποίος εισέρχεται τελείως στο δίσκο του πρωτεύοντος αστέρα κατά την πρώτη έκλειψη και που επισκιάζεται πλήρως κατά τη δεύτερη.



Σχήμα 4.4: Διορθωμένη καμπύλη φωτός για τον τύπου W UMa διπλό εκλειπτικό KID 10618251. Η περίοδος της αναδίπλωσης διπλασιάστηκε για να απεικονιστούν και οι δύο εκλείψεις. Η ομοιότητα με ημιτονοειδή καμπύλη οφείλεται στην εγγύτητα των μελών του συστήματος.

5

Συμπεράσματα

Επανερχόμενοι στους στόχους που τέθηκαν στην § 1.6, θα ερευνήσουμε αν αυτοί εκπληρώθηκαν:

1. Πραγματοποίηση των αλγορίθμων και έλεγχος καλής λειτουργίας τους

Οι αλγόριθμοι προγραμματίστηκαν και πέρασαν μία σειρά ελέγχων που είχαν σκοπό να επιβεβαιώσουν τη σωστή έξοδό τους. Χρησιμοποιήθηκαν μαθηματικά πακέτα και υπολογισμοί στο χέρι για μικρές εισόδους.

2. Είναι τόσο χρονοβόροι όσο περιμέναμε;

Οι αλγόριθμοι εκτελέστηκαν για διάφορα μεγέθη εισόδου. Η υπολογιστική πολυπλοκότητα επιβεβαιώθηκε και οι χρονομετρήσεις ανέδειξαν τη σημασία της χρήσης γρήγορων αλγορίθμων στις αστρονομικές έρευνες ευρέως πεδίου.

3. Ποια είναι η επιτυχία τους στην εύρεση περιοδικοτήτων σε χρονοσειρές με διάφορες παραμέτρους: σχήμα, περίοδος, επίπεδα θορύβου;

Οι αλγόριθμοι έτρεξαν συνολικά για 10000 καμπύλες φωτός που δημιουργήθηκαν από εμάς. Για ημιτονοειδή σχήματα ο FFT είχε καλύτερα αποτελέσματα όπως υποθέταμε. Για μικρούς ορθογώνιους παλμούς, ενώ αναμέναμε τον FFA να τα πάει καλύτερα από τον FFT, τα τεστ δεν επιβεβαίωσαν αυτό το γεγονός. Υποθέτουμε ότι αυτό οφείλεται σε κακή ερμηνεία των εξόδων των αλγορίθμων παρά σε ανακριβείς εξόδους.

4. Μπορούν να χρησιμοποιηθούν στην ανίχνευση περιόδων των τριών αντιπροσωπευτικών τάξεων αντικειμένων (Κηφίδες, διπλοί εκλειπτικοί και διαβάσεις εξωπλανητών);

Παρά την απλοϊκή προσέγγιση για την εύρεση περιοδικοτήτων, ο συνδυασμός των δύο αλγορίθμων κατάφερε να εντοπίσει τις περιόδους των δώδεκα περιπτώσεων με καλή ακρίβεια. Μόνο στην περίπτωση του FFA και των διαβαίνοντων εξωπλανητών υπήρξαν αποτυχίες, γεγονός που συζητήσαμε αναλυτικά στην § 4.3. Επίσης μελετήσαμε τις επιδράσεις των σφαλμάτων των εκτιμήσεων στα αναδιπλωμένα προφίλ (επιμήκυνση, εξομάλυνση κτλ.) καθώς και του εντοπισμού εσφαλμένων αρμονικών.

Μία πιο προσεκτική εκμετάλλευση των εξόδων των δύο αλγορίθμων μπορεί να αναδείκνυε την υπεροχή του FFA απέναντι στον FFT για μικρού κύκλου εργασιών παλμούς (Staelin, 1969; Burns & Clark, 1969; Kondratiev et al., 2009), κάτι που πιθανώς θα ήταν χρήσιμο στην ανακάλυψη διαβαίνοντων εξωπλανητών.

6

Προτάσεις

Η εμπειρία μας έδειξε κάποια σοβαρά προβλήματα αλλά και προκάλεσε τον ενδιαφέρον του γράφοντος για μια πιο προσεκτική προσέγγιση στο πρόβλημα της αναζήτησης περιοδικοτήτων με τον αλγόριθμο FFA και για την ανάπτυξη ενός πιο ολοκληρωμένου λογισμικού.

6.1 Αλλαγές

6.1.1 Συμβάσεις προγραμματισμού

Η επιλογή κινητής υποδιαστολής τιμών για τα στοιχεία των καμπύλων φωτός είναι χρονοβόρα. Θα μπορούσε να αυξηθεί η ταχύτητα των αλγορίθμων χρησιμοποιώντας μία λέξη υπολογιστή για την αποθήκευση δύο ακέραιων τιμών. Προσεκτική κανονικοποίηση απαιτείτε για να αποφευχθεί ανεπιθύμητη εναλλαγή προ-

σήμων.

6.1.2 Κατάλληλα στατιστικά μέτρα σημαντικότητας

Ακόμα και μία τέλεια έξοδος ενός αλγορίθμου μπορεί να καταστεί άχρηστη αν χρησιμοποιηθούν ακατάλληλες στατιστικές μέθοδοι. Μία μέθοδος για τη σωστή εκτίμηση του λόγου σήματος – θορύβου θα έπρεπε να αναληφθεί (ίσως window functions για τον FFA και σωστή ανάλυση των αρμονικών για τον FFT).

6.1.3 Τακτοποίηση

Η μέθοδός μας για την μετατροπή άνισων στα διαστήματα χρονοσειρών σε κλασικές χρονοσειρές εισάγει σφάλματα στα phase bins και δουλεύει μόνο αν η δειγματοληψία είναι σχεδόν σταθερή, αποτρέποντας την εκμετάλλευση δεδομένων με μεγάλα κενά. Πιθανές λύσεις:

- Μέθοδος παρεμβολής
- Παρεμβολή μόνο για τις περιοχές με κενά
- Re-binning των χρονοσειρών σε μεγαλύτερα διαστήματα, ρισκάροντας μεν την απώλεια πληροφορίας, αυξάνοντας την ταχύτητα της επεξεργασίας δε.

6.2 Μελλοντικές προεκτάσεις

6.2.1 Μελέτη τάσεων και έγχρωμων θορύβων

Η επίδραση τάσεων (trends) και διαφόρων τύπων θορύβου θα πρέπει να μελετηθούν. Η Γκαουσιανή κατανομή του θορύβου είναι μία προσέγγιση. Στις περισσότερες φυσικές διεργασίες καταμέτρησης, συναντούμε την κατανομή Poisson.

6.2.2 Συνεργασία FFT/FFA/LS

Προτείνουμε μία γρήγορη μέθοδο για αυτόματη αναζήτηση περιοδικοτήτων σε καμπύλες φωτός:

1. Αφαίρεση τάσεων και μείωση θορύβου με φίλτρα
2. Re-binning και μέθοδος παρεμβολής για μετατροπή σε χρονοσειρά σταθερής δειγματοληψίας
3. FFT και εντοπισμός ενός πλήθους περιοχών ενδιαφέροντος με πιθανές περιοδικότητες, εκμεταλλευόμενη την ταχύτητα ($O(n \log n)$).
4. Εκτέλεση του FFA για τις στενές αυτές περιοχές (η πολυπλοκότητα $O(n^2)$) δε θα εισάγει σημαντικές καθυστερήσεις) και νέα εκτίμηση πιθανών περιόδων με μεγαλύτερη ακρίβεια
5. Χρήση του περιοδογράμματος Lomb–Scargle που αναλύει μη σταθερής δειγματοληψίας χρονοσειρές, πάνω στην αρχική μη re-binned μορφή στις πολύ στενές περιοχές που έδειξε ο FFA.

Μαθηματικοί ορισμοί και αποδείξεις

A'.1 Big O notation

Ο ακριβής χρόνος εκτέλεσης ενός αλγορίθμου, $f(n)$, εξαρτάται από το μέγεθος της εισόδου, n , και στις περισσότερες των περιπτώσεων πρόκειται για μία πολύπλοκη έκφραση. Επειδή συνήθως ενδιαφερόμαστε για την αποτελεσματικότητα ενός αλγορίθμου όταν η είσοδος είναι μεγάλη (οπότε η ταχύτητα είναι σημαντική), εφαρμόζουμε την ασυμπτωτική ανάλυση.

Αυτό το κάνουμε εστιάζοντας στον όρο μεγαλύτερης τάξης της έκφρασης, απορρίπτοντας τον συντελεστή του (τη σταθερά). Για παράδειγμα, η έκφραση $\frac{2}{\log 2} n^2 \log n + 3.75n^2 + 28$ γίνεται $n^2 \log n$.

Στο παρόν κείμενο εμφανίζεται ο συμβολισμός $O(f'(n))$ που αναφέρεται σε αυτήν τη διαδικασία. Ο μαθηματικός ορισμός του λεγόμενου 'big O notation' είναι:

$$f(n) = O(f'(n)) \text{ as } n \rightarrow \infty$$

αν και μόνο αν υπάρχουν μία θετική σταθερά M και ένα μεγάλο n_0 ώστε

$$|f(n)| \leq M|f'(n)| \text{ για όλα τα } n \geq n_0$$

Α΄.2 Υπολογιστική πολυπλοκότητα για ‘τυφλές’ έρευνες με FFA

- Η πολυπλοκότητα του FFA σε μία είσοδο μεγέθους N για βασική περίοδο P είναι ανάλογη του $N \log_2 \left\lfloor \frac{N}{P} \right\rfloor$
- Στο σενάριο χειρότερης περίπτωσης, εφαρμόζουμε των αλγόριθμων για εύρος περιόδων $[2, N/2]$
- Αυτό οδηγεί σε ολική πολυπλοκότητα $O(N^2)$

Απόδειξη

$$\begin{aligned}
 & \sum_{P=2}^{N/2} \left(N \log_2 \left\lfloor \frac{N}{P} \right\rfloor \right) \simeq \\
 & \simeq \sum_{P=2}^{N/2} \left[N \log_2 \left(\frac{N}{P} \right) \right] \simeq \\
 & \simeq \int_2^{N/2} N \log_2 \frac{N}{P} dP = \\
 & = \frac{1}{\ln 2} \int_2^{N/2} N \ln \frac{N}{P} dP = \\
 & = \frac{1}{\ln 2} \int_2^{N/2} (N \ln N - N \ln P) dP = \\
 & = \frac{1}{\ln 2} [(N \ln N) P - N (P \ln P - P)]_{P=2}^{P=N/2} =
 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\ln 2} \left[\frac{N^2}{2} - 2\ln 2 \cdot N - 2(1 - \ln 2) N \right] = \\
&= O(N^2)
\end{aligned}$$

Α΄.3 Τύποι τακτοποίησης

Σύνοψη

- t_i
- Αναμενόμενες τιμές: $\tau_i = \tau_0 + i \cdot p_s$
- Μέτρο καλής προσαρμογής: $g = \sum_{i=0}^{N-1} (t_i - \tau_i)^2$
- Προσδιορισμός των τ_0 και p λύνοντας $\frac{\partial g}{\partial \tau_0} = 0$ και $\frac{\partial g}{\partial p_s} = 0$

Επίλυση

$$\begin{aligned}
&\frac{\partial g}{\partial \tau_0} = 0 & \frac{\partial g}{\partial p_s} = 0 \\
&\frac{\partial}{\partial \tau_0} \sum_{i=0}^{N-1} (t_i - \tau_i)^2 = 0 & \frac{\partial}{\partial p_s} \sum_{i=0}^{N-1} (t_i - \tau_i)^2 = 0 \\
&\sum_{i=0}^{N-1} \left[\frac{\partial}{\partial \tau_0} (t_i - \tau_0 - i \cdot p_s)^2 \right] = 0 & \sum_{i=0}^{N-1} \left[\frac{\partial}{\partial p_s} (t_i - \tau_0 - i \cdot p_s)^2 \right] = 0 \\
&\sum_{i=0}^{N-1} [(t_i - \tau_0 - i \cdot p_s) (-2)] = 0 & \sum_{i=0}^{N-1} [(t_i - \tau_0 - i \cdot p_s) (-2i)] = 0 \\
&\sum_{i=0}^{N-1} t_i - \sum_{i=0}^{N-1} \tau_0 - p_s \sum_{i=0}^{N-1} i = 0 & \sum_{i=0}^{N-1} (i \cdot t_i) - \tau_0 \sum_{i=0}^{N-1} i - p_s \sum_{i=0}^{N-1} i^2 = 0
\end{aligned}$$

Χρησιμοποιώντας ότι

$$\sum_{i=0}^{N-1} i = \frac{N(N-1)}{2}$$

$$\sum_{i=0}^{N-1} i^2 = \frac{N(N-1)(2N-1)}{6}$$

καταλήγουμε στο σύστημα εξισώσεων

$$N\tau_0 + \frac{N(N-1)}{2}p_s = \sum_{i=0}^{N-1} t_i$$

$$\frac{N(N-1)}{2}\tau_0 + \frac{N(N-1)(2N-1)}{6}p_s = \sum_{i=0}^{N-1} (i \cdot t_i)$$

Τέλος, υπάρχει μία μοναδική λύση:

$$\tau_0 = \frac{2}{N^2 + N} \left[(2N-1) \sum_{i=0}^{N-1} t_i - 3 \sum_{i=0}^{N-1} (i \cdot t_i) \right]$$

$$p_s = \frac{6}{N(N^2 - 1)} \left[2 \sum_{i=0}^{N-1} (i \cdot t_i) - (N-1) \sum_{i=0}^{N-1} t_i \right]$$

Γεννήτριες τυχαίων αριθμών

Οι τεχνητές καμπύλες φωτός που χρησιμοποιήθηκαν για να ελεγχθούν οι αλγόριθμοι FFA και FFT έπρεπε να ‘μολυνθούν’ με Γκαουσιανό θόρυβο (βλέπε § 1.2). Η παραγωγή τέτοιων τυχαίων αριθμών απαιτεί τις παρακάτω συναρτήσεις. Σημειώστε ότι η χρήση των ενσωματωμένων συναρτήσεων (όπως η `rand()`) αλλά τους σκοπούς της παρούσας εργασίας, η χρήση πολυπλοκότερων μεθόδων δε θεωρήθηκε αναγκαία (Press et al., 2007).

B'.1 Ομοιόμορφη κατανομή

`rnd_uniform()`

Η παραγωγή τυχαίων αριθμών τυπικής ομοιόμορφης κατανομής¹ είναι τριμμένη καθώς το ‘αρχείο κεφαλής’ της C, `stdlib.h` περιλαμβάνει τη συνάρ-

¹ Σταθερή πυκνότητα πιθανότητας στο διάστημα $[0,1]$ και μηδέν οπουδήποτε αλλού

τηση `rand()` που επιστρέφει έναν ακέραιο στο διάστημα $[0, \text{RAND_MAX}]$. Η σταθερά `RAND_MAX` ορίζεται στο ίδιο αρχείο συνήθως ισούται με 32767.

Έτσι, διαιρούμε την επιστρεφόμενη τιμή με το `RAND_MAX` και πολλαπλασιάζουμε με 1.0 για να αποτρέψουμε την ακέραια διαίρεση (που θα έδινε αποτέλεσμα 0.0).

Η ανάλυση (32768 διαφορετικά επίπεδα) είναι περισσότερο από αρκετή στα πλαίσια των ελέγχων που θα πραγματοποιήσουμε εδώ.

```
1 float rnd_uniform ()
2 {
3     return rand () * 1.0 / RAND_MAX;
4 }
```

Listing B'.1: Μία γεννήτρια τυχαίων αριθμών τυπικής ομοιόμορφης κατανομής

B'.2 Κανονική κατανομή

`rnd_gaussian()`

Εδώ βρίσκεται μία πραγματοποίηση της μεθόδου Marsaglia (Marsaglia & Bray, 1964) για την παραγωγή τυχαίων αριθμών που ανήκουν στην τυπική κανονική κατανομή¹. Η συνάρτηση χρησιμοποιεί την `rnd_uniform()` που περιγράφηκε παραπάνω.

```
1 float rnd_gaussian ()
2 {
3     float R=0,a,b;
4     do
5     {
6         a=rnd_uniform () * 2 - 1;
7         b=rnd_uniform () * 2 - 1;
8         R=a*a+b*b;
9     } while (R>=1 || R==0);
10
11     return a * sqrt(-2 * log (R) / R);
12 }
```

Listing B'.2: Μία γεννήτρια τυχαίων αριθμών τυπικής κανονικής κατανομής

¹ μηδενικού μέσου όρου 0 και τυπικής απόκλισης 1.0

“Πάντα να προγραμματίζεις σαν ο άνθρωπος
που θα διατηρεί τον κώδικά σου να είναι ένας
βίαιος ψυχοπαθής που γνωρίζει πού μένεις.”

John F. Woods

Κύριος κώδικας

Γ.1 Τύποι μεταβλητών

```
1 typedef double TIME; // type for time in BJD
2 typedef float BIN; // type for measurement in arbitrary units
3
4 typedef struct
5 {
6     int KID; // Kepler ID
7     TIME ASASperiod; // All Sky Automated Survey Period
8     char *notes; // NOTES
9 } STAR_INFO;
10
11 typedef struct
12 {
13     BIN *bins; // list of measurements
14     int size; // size of tables
15     TIME *times; // list of observation times in days
```

```

16     TIME timeref;      // time reference in BJD
17 } TIME_BIN_TABLE;
18
19 typedef struct
20 {
21     BIN *bins;          // table of measurements
22     int size;           // size of table
23     TIME timeref;       // time reference in BJD
24     TIME timeerr;       // regularization time error
25     TIME interval;      // sampling rate
26 } BIN_TABLE;
27
28 typedef struct
29 {
30     float period;        // detected period
31     float dp;            // error in period
32     float significance;  // significance
33     int outcome;         // _SUCCESS or _FAILURE
34 } FFA_RESULT;

```

Listing Γ'.1: Various newly defined data types

Γ'.2 Ρουτίνα τακτοποίησης

```

1 int REGULARIZE(const TIME_BIN_TABLE *src, BIN_TABLE *dst)
2 {
3     printf("\nREGULARISE REPORT\n\n");
4     int n=(*src).size;
5     (*dst).size=n;
6     (*dst).bins=malloc(n*sizeof(BIN));
7
8     int i;
9     TIME ti=0,sumti=0,sumiti=0;
10    for(i=0;i<n;i++)
11    {
12        (*dst).bins[i]=(*src).bins[i];    // copy values
13    }

```

```

14     ti=(*src).times[i];           // but also do the
15     sumti+=ti;                     // ... statistics for...
16     sumiti+=i*ti;                  // ... the regularization.
17 }
18
19 TIME delta_tau=(2.0/(1.0*n-1)*sumiti - sumti)*6/(n*(n+1.0));
20 TIME tau=((2.0*n-1)/3)*sumti-sumiti)*6/(n*(n+1.0));
21 TIME error=0;
22 for(i=0;i<n;i++)
23 {
24     ti=(*src).times[i]-tau-i*delta_tau;
25     ti=ti*ti;
26     error+=ti;
27 }
28 error=sqrt(error/n);
29
30 (*dst).timeref=(*src).timeref+tau;
31 (*dst).interval=delta_tau;
32 (*dst).timeerr=error;
33
34 printf("New time offset \t%lf BJD\n",(*dst).timeref);
35 printf("Sampling rate \t\t%lf min\n",(*dst).interval*24*60);
36 printf("Regularisation error \t%lf min\n",(*dst).timeerr*24*60);
37
38 return _SUCCESS;
39 }

```

Listing Γ'.2: The implementation of FFT algorithm

Γ'.3 Κώδικας FFT

```

1 void RECURSIVE_FFT(complex *input,complex *output,int N,int step)
2 {
3     int i;
4     if (step<N)
5     {
6         RECURSIVE_FFT(output,input,N,step*2);

```

```

7     RECURSIVE_FFT(output+step , input+step ,N, step*2);
8
9     for ( i=0; i<N; i+=2*step )
10    {
11        complex tmp = cexp(-I*M_PI*i/N)*output[i+step];
12        input[i/2] = output[i] + tmp;
13        input[(i+N)/2]=output[i] - tmp;
14    }
15 }
16 }
17
18 float FFT(const BIN_TABLE *data)
19 {
20     printf("\nFFT REPORT\n-----\n");
21     int n=(int) pow(2,(int) log2((*data).size));
22
23     complex *input=malloc(n*sizeof(complex));
24     complex *output=malloc(n*sizeof(complex));
25
26     int i;
27     for ( i=0; i<n; i++)
28     {
29         input[i]=(complex) (*data).bins[i];
30         output[i]=(complex) (*data).bins[i];
31     }
32
33     RECURSIVE_FFT(input , output ,n,1);
34
35     // Save power spectrum
36     FILE *fftfile=fopen("fftspectrum.txt","w");
37     if (fftfile==NULL)
38     {
39         printf("\n\nError opening file\n\n");
40         return _FAILURE;
41     }
42     printf("> Spectrum file created.\n\n");
43     fprintf(fftfile , "Frequency , Magnitude\n");
44
45     float max=0;

```

```

46  float x;
47  int imax=0;
48
49  for ( i=1; i<n/2; i++)
50  {
51      x=cabs ( input [ i ] );
52      if (i<FFToutput) fprintf ( fftfile , "%f,%f\n" , i / (( * data ) . interval
53          *(1.0*n)) , (x) );
54      if ( x>max )
55      {
56          max=x;
57          imax=i;
58      }
59  }
60  fclose ( fftfile );
61
62  float f=imax;
63  float pmin=( * data ) . interval *n / ( f+0.5 );
64  float pmax=( * data ) . interval *n / ( f-0.5 );
65  float period=(pmax+pmin) / 2;
66  float dp=(pmax-pmin) / 2;
67
68  printf ( "Frequency: \t\t%d bins\n" , imax );
69  printf ( "Period: \t\t%f days \t+/- %f\n" , period , dp );
70
71  return period;
72 }

```

Listing Γ'.3: Ο κώδικας του FFT

Γ'.4 Κώδικας FFA

```

1  FFA_RESULT FFA ( BIN *input , int bins , int P )
2  {
3      FFA_RESULT result;
4

```

```

5 // dimensions , blocks , etc .
6 int M=(int) bins/P;
7 int steps=(int) log2(M);
8 M=(int) pow(2,steps);
9 int N=M*P;
10
11
12 // allocating memory for temporary arrays
13 BIN *A=malloc(N*sizeof(BIN));
14 BIN *B=malloc(N*sizeof(BIN));
15
16 // initializing
17 int i;
18 for(i=0;i<N;i++)
19 {
20     A[i]=input[i];
21     B[i]=0;
22 }
23
24 // FFA
25 int step,k,shift;
26 for(step=1;step<=steps;step++)
27 {
28     int blocklen=(int) pow(2,step);
29     int blocks=(int) M/blocklen;
30     int block;
31     for(block=0;block<blocks;block++)
32     {
33         int row;
34         for(row=0;row<blocklen;row++)
35         {
36             for(k=0;k<P;k++)
37             {
38                 shift=(int) (row+1)/2;
39                 B[(block*blocklen+row)*P+k]=
40                 A[(block*blocklen+((int) row/2))*P+k]+
41                 A[(block*blocklen+((int) (row+blocklen)/2))*P+(k+
42                     shift)%P];

```



```

43     }
44 }
45     for ( i=0; i<N; i++) A[ i ]=B[ i ];
46 }
47
48 // period and significance estimate
49
50 float max=0;
51 float period_at_max=0;
52
53 int fold;
54 for ( fold=0; fold<M; fold++)
55 {
56     float meanx=0;
57     float meanx2=0;
58     for ( i=0; i<P; i++)
59     {
60         BIN x=A[ fold *P+i ];
61         meanx+=x;
62         meanx2+=x*x;
63     }
64     meanx=meanx /P;
65     meanx2=meanx2 /P;
66     float SD=sqrt ( P/( P-1.0) ) * sqrt ( meanx2-meanx*meanx );
67     if (SD>max)
68     {
69         max=SD;
70         period_at_max=P+( fold /(1.0*M-1.0));
71     }
72 }
73 free (A);
74 free (B);
75
76 result.outcome=_SUCCESS;
77 result.significance=max/ sqrt (M);
78 result.dp=1/(2.0*(M-1));
79 result.period=period_at_max;
80
81 return result;

```

```

82 }
83
84 FFA_RESULT FFAsearch(BIN_TABLE *data,int periodmin,int periodmax)
85 {
86     int bins=(*data).size;
87     printf("\nFFA REPORT\n\n");
88     BIN *input=malloc(bins*sizeof(BIN));
89     int i;
90     for(i=0;i<bins;i++)
91     {
92         input[i]=(*data).bins[i];
93     }
94
95     int period,count=0;
96     FFA_RESULT result;
97     result.period=0;
98     result.significance=0;
99     result.outcome=_FAILURE;
100
101     FILE *ffafile=fopen("ffaperiodogram.txt","w");
102     if(ffafile==NULL)
103     {
104         printf("\n\nError opening file\n\n");
105         return result;
106     }
107     fprintf(ffafile,"Period,Significance\n");
108     printf("> Periodogram file created.\n\n");
109
110     BIN max=0;
111     float period_at_max=0,dp_at_max,x,meanx=0,meanx2=0;
112
113     for(period=periodmin;period<=periodmax;period++)
114     {
115         count++;
116
117         result=FFA(input,bins,period);
118         x=result.significance;
119         fprintf(ffafile,"%f,%f\n",(*data).interval*result.period,
            result.significance);

```

```

120
121     meanx+=x;
122     meanx2+=x*x;
123
124     if (x>max)
125     {
126         period_at_max=result.period;
127         dp_at_max=result.dp;
128         max=x;
129     }
130 }
131
132 meanx=meanx/count;
133 meanx2=meanx2/count;
134 float stdev = sqrt(count/(1.0*count-1.0))*sqrt(meanx2-meanx*
        meanx);
135 result.significance=(max-meanx)/stdev;
136 result.period=period_at_max>(*data).interval;
137 result.dp=dp_at_max>(*data).interval;
138 result.outcome=_SUCCESS;
139
140 printf("Period: \t\t%f days \t+/- %f\n",result.period,result.dp)
        ;
141
142 fclose(ffafile);
143 return result;
144 }

```

Listing Γ'.4: Ο κώδικας του FFA

Γ'.5 Ρουτίνα αναδίπλωσης

```

1 int FOLD(TIME_BIN_TABLE *data, TIME p, char *path)
2 {
3     FILE *file=fopen(path, "w");
4     fprintf(file, "phase, magnitude\n");
5

```

```

6   int n=(*data).size;
7   TIME start=(*data).times[0];
8   TIME end=(*data).times[n];
9
10  BIN b,max=0;
11  TIME t,timeatmax=0;
12  int i;
13  for (i=0;i<n;i++)
14  {
15      b=fabs((*data).bins[i]);
16      t=(*data).times[i];
17
18      if (t<start) start=t;
19      if (t>end) end=t;
20
21      if (b>max)
22      {
23          max=b;
24          timeatmax=t;
25      }
26  }
27
28  for (i=0;i<n;i++)
29  {
30      t=(*data).times[i]-0.5*timeatmax; // start;
31      t=(t-trunc(t/p)*p)/p-0.5;
32      fprintf(file,"%lf,%f\n",t,(*data).bins[i]/max);
33  }
34
35  printf("\n> Folding at p=%lf saved to: %s\n",p,path);
36  fclose(file);
37 }

```

Listing Γ'.5: Η ρουτίνα αναδίπλωσης

Γ'.6 Ρουτίνα Re-binned αναδίπλωσης

```

1  int FOLD_REBINNED(TIME_BIN_TABLE *data ,int REBINNING, TIME p,char *
    path)
2  {
3      int i;
4      FILE *file=fopen(path,"w");
5      fprintf(file,"phase bin,magnitude\n");
6
7      int n>(*data).size;
8      BIN *bins=malloc(REBINNING*sizeof(BIN));
9      int *bincount=malloc(REBINNING*sizeof(int));
10     for (i=0;i<REBINNING;i++)
11     {
12         bins[i]=0;
13         bincount[i]=0;
14     }
15
16     TIME start=(*data).times[0];
17     TIME end=(*data).times[n-1];
18
19     BIN b;
20     TIME t,timeatmax=0;
21
22     for (i=0;i<n;i++)
23     {
24         TIME t=(*data).times[i]-start;
25         t=t/p-trunc(t/p);
26         int pos=(int)(t*REBINNING);
27         if (pos>=REBINNING) pos=REBINNING-1;
28         bins[pos]+=(*data).bins[i];
29         bincount[pos]++;
30     }
31
32     BIN max=0;
33     for (i=0;i<REBINNING;i++)
34     {
35         int norm=bincount[i];
36         if (norm!=0) bins[i]/=norm;
37         if (fabs(bins[i])>=max) max=fabs(bins[i]);
38     }

```

```
39
40     for ( i=0;i<REBINNING;i++)
41     {
42         int norm=bincount[i];
43         if (norm!=0) fprintf( file ,"%d,%f\n",i+1,bins[i]/max);
44     }
45
46     printf("\n> Rebinned (%d) folding at p=%lf saved to: %s\n",
47           REBINNING,p,path);
47     fprintf(outputfile,"\n> Rebinned (%d) folding at p=%lf saved to:
48           %s\n",REBINNING,p,path);
48     fclose( file );
49 }
50 }
```

Listing Γ'.6: Η ρουτίνα rebinned αναδίπλωσης

Ευρετήριο

- All Sky Automated Survey, 43, 44, 59, 88, 89
- DC offset, 18, 42
- FFA, i, 2, 11, 13–15, 32, 37, 40, 41, 43, 46, 47, 59, 87, 89, 90, 103
- FFT, i, 2, 11, 16, 21, 25, 32, 40, 41, 43, 46, 47, 59, 86–89, 103
- Peak period, 10, 14
- Αντίστροφος DFT, 18
- Αποστολή Kepler, 32, 43–45, 58, 59, 88
- Γκαουσιανός θόρυβος, 4, 8, 88, 96, 103
- Διάγραμμα ροής, 11, 12, 23, 24
- Διακριτός μετασχηματισμός Fourier, 18, 20, 22, 36
- Διακριτός μετασχηματισμός Fourier, 22
- Θεώρημα δειγματοληψίας, 18
- Κεντρικό οριακό θεώρημα, 4, 88
- Κύκλος εργασίας, 41, 42, 87, 89
- Λόγος σήματος θορύβου, 4, 5, 7, 8, 38, 87, 88, 96
- Μέθοδος ελαχίστων τετραγώνων, 46, 101
- Μετασχηματισμός, 16
- Μη τακτικές χρονοσειρές, 2
- Νόμος των μεγάλων αριθμών, 8
- Ρυθμός Nyquist, 18
- Σειρά Fourier, 16
- Συνεχής μετασχηματισμός Fourier, 17
- Συχνότητα Nyquist, 19
- Τακτικές χρονοσειρές, 2
- Υπολογιστική πολυπλοκότητα, 10, 13, 14, 18, 22, 40, 41, 49, 85, 86, 93, 99, 100
- Φασματική διάχυση, 20, 21, 37, 87
- Χρονοσειρά, 2, 3

Βιβλιογραφία

- Borucki, W. J. et al. (2011). Characteristics of planetary candidates observed by kepler. ii. analysis of the first four months of data. *The Astrophysical Journal*, 736(1), 19.
- Burns, W. R. & Clark, B. G. (1969). Pulsar search techniques. *Astronomy & Astrophysics*, 2, 280–287.
- Christiansen, J. L., Van Cleve, J. E., Jenkins, J. M., Caldwell, D. A., Allen, C. L., Barclay, T., Bryson, S. T., Burke, C. J., Clarke, B. D., Cote, M. T., Fanelli, M. N., Gilliland, R. L., Girouard, F., Haas, M. R., Hall, J. R., Ibrahim, K., Kinemuchi, K., Klaus, T. C., Kolodziejczak, J. J., Li, J., Machalek, P., & McCauliff, S. D. (2011). Kepler data release 11 notes. *KSCI–19051–001*.
- Chu, E. (2008). *Discrete and Continuous Fourier Transforms: Analysis, Applications and Fast Algorithms*. CRC Press.
- Cooley, J. W. & Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19, 297–301.
- Cowpertwait, S. P. P. & Metcalfe, V. A. *Introductory Time Series with R*. Springer.
- Heideman, M. T., Johnson, D. H., & Burrus, C. S. (1985). Gauss and the history of the fast fourier transform. *Archive for History of Exact Sciences*, 34(3), 265–267.
- Herrero, E., Morales, J. C., Ribas, I., & Naves, R. (2011). Wasp–33: The first δ scuti exoplanet host star. *Astronomy and Astrophysics Letters*, 526(L10), 4.
- Hipel, K. W. & McLeod, A. I. (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier.

- Horne, K. (2012). As5001: Advanced (astronomical) data analysis. Lecture Notes.
- Kondratiev, V. I., McLaughlin, M. A., Lorimer, D. R., Burgay, M., Possenti, A., Turolla, R., Popov, S. B., & Zane, S. (2009). New limits on radio emission from x-ray dim isolated neutron stars. *The Astrophysical Journal*, 702, 692–706.
- Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39, 447–462.
- Marsaglia, G. & Bray, T. A. (1964). A convenient method for generating normal variables. *Society for Industrial and Applied Mathematics Review*, 6(3), 260–264.
- Mayor, M. & Queloz, D. (1995). A jupiter-mass companion to a solar-type star. *Nature*, 378, 355–359.
- O'Donovan, F. T., Charbonneau, D., Mandushev, G., Dunham, E. W., Latham, D. W., Torres, G., Sozzetti, A., Brown, T. M., Trauger, J. T., Belmonte, J. A., Rabus, M., Almenara, J. M., Alonso, R., Deeg, H. J., Esquerdo, G. A., Falco, E. E., Hillenbrand, L. A., Roussanova, A., Stefanik, R. P., & Winn, J. N. (2006). Tres-2: The first transiting planet in the kepler field. *The Astrophysical Journal*, 651(1), L61–L64.
- Percy, J. R. (2007). *Understanding Variable Stars*. Cambridge University Press.
- Perryman, M. (2011). *The Exoplanet Handbook*. Cambridge University Press.
- Pigulski, A., Pojmański, G., Pilecki, B., & Szczygieł, D. M. (2009). The all sky automated survey. the catalog of variable stars in the kepler field of view. *Acta Astronomica*, 59(1), 33–46.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press.
- Sabin, W. E. (2008). *Discrete-Signal Analysis and Design*. Wiley-Interscience.
- Scargle, J. D. (1982). Studies in astronomical time series analysis. ii. statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263, 835–853.

- Shporer, A., Jenkins, J. M., Rowe, J. F., Sanderfer, D. T., Seader, S. E., Smith, J. C., Still, M. D., Thompson, S. E., Twicken, J. D., & Welsh, W. F. (2011). Detection of koi-13.01 using the photometric orbit. *The Astronomical Journal*, 142(6), 7.
- Shu, H. F. (1982). *Physical Universe: An Introduction to Astronomy*. University Science Books.
- Spiegel, M. R. (1975). *Shaum's Outline of Probability and Statistics*. McGraw-Hill.
- Staelin, D. H. (1969). Fast folding algorithm for detection of periodic pulse trains. *Proceedings of the IEEE*, 57(4), 724–725.
- Thomson, S. E., Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., Barclay, T., Bryson, S. T., Burke, C. J., Clarke, B. D., Girouard, F., Haas, M. R., Hall, J. R., Ibrahim, K., Klaus, T. C., Kolodziejczak, J. J., Li, J., & McCauliff, S. D. (2013). Kepler data release 19 notes. *KSCI-19059-001*.
- Wall, J. V. & Jenkins, C. R. (2003). *Practical Statistics for Astronomers*. Cambridge University Press.
- Warner, B. D. (2006). *Lightcurve Photometry and Analysis*. Springer.
- Welsh, W. F., Orosz, J. A., Seager, S., Fortney, J. J., Jenkins, J., Rowe, J. F., Koch, D., & Borucki, W. J. (2010). The discovery of ellipsoidal variations in the kepler light curve of hat-p-7. *The Astrophysical Journal Letters*, 713(2), L145–L149.
- Zakai, A. (2009). The rise of modern science and the decline of theology as the ‘queen of sciences’ in the early modern era. *Renaissance and Reformation Review*, 9(2), 125–151.

