

# Data mining in astronomy: classification of eclipsing binaries.

Oleg Malkov

*Institute of Astronomy, Russian Academy of Sciences;  
Physics Department, Moscow State University*

Belgrade, Jul 1, 2010

# Plan

- Binary stars
- Eclipsing binary stars
- Modern catalogues and lists of eclipsing binary stars
- Classification of eclipsing binary stars
- Determination of fundamental stellar parameters and evolutionary status of close binary systems
- Astrophysical applications

# Binary stars

Belgrade, Jul 1, 2010

# Why study binaries?

- They are very common: perhaps 90% of stellar systems are binaries.
- There are still many puzzles to solve about their structure and evolution.
- They provide accurate data that are very difficult to determine for single stars such as masses and radii.

The direct determination of the mass of any astronomical object requires measurable gravitational interaction between at least two objects.

# Gravitational interactions

- galaxy–galaxy interaction: the distances and separations are so large that no detectable motion on the plane of the sky is possible
- star–planet interaction: the objects contrast so greatly in brightness that (outside the solar system) only the highest possible precision can resolve the objects; in such cases only the star's motions are detectable, and the properties of that star must be assumed in order to deduce the properties of the planet

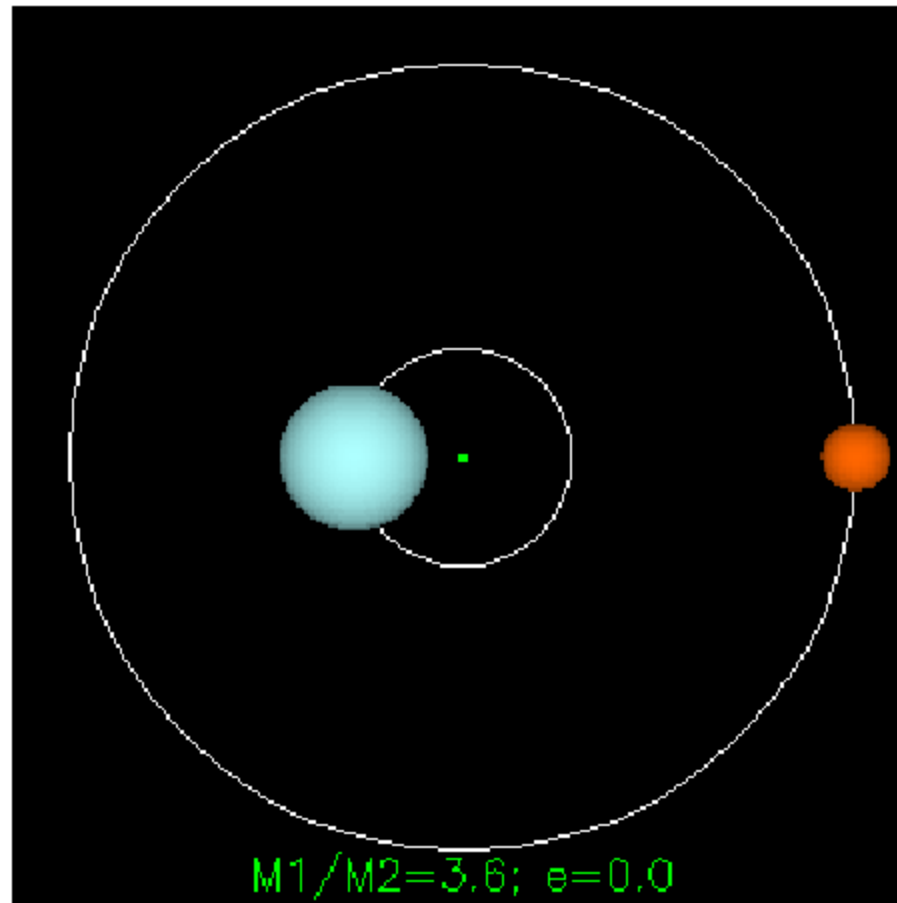
# Gravitational interactions

- planet–satellite interaction: solar system only
- star–star interaction: the variations in position and velocity caused by orbital motion are detectable for a wide range of stellar separation and at least a factor of 5 in brightness (corresponds to a magnitude difference of about 1.<sup>m</sup>75)

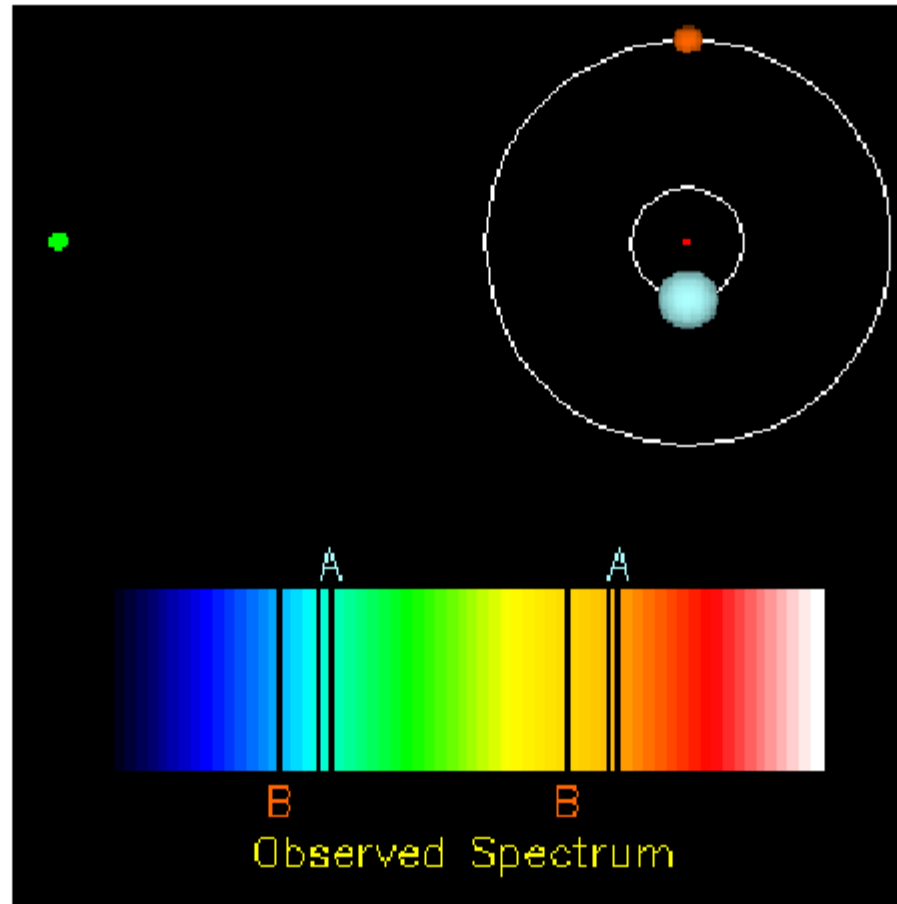
# Types of binaries

- Described by the observational technique. A system is classified depending, mostly, on its semi-major axis and distance.

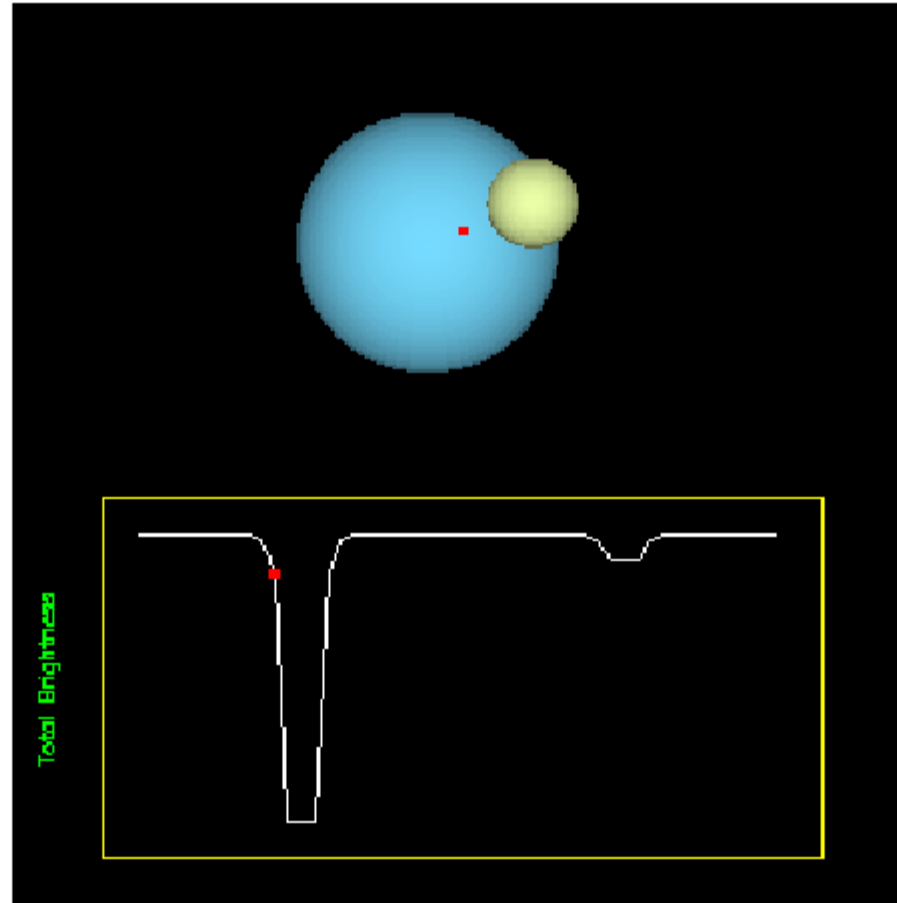
# Visual binaries



# Spectroscopic binaries



# Eclipsing (photometric) binaries

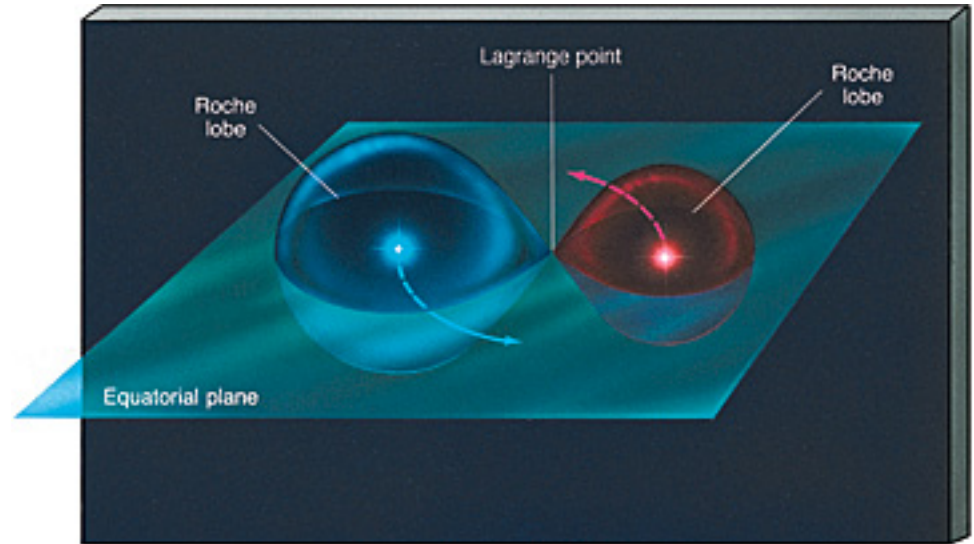


# Types of binaries: Kopal scheme

- Described by stability within equipotential surfaces (Roche lobes).

# Assumptions of the Standard Roche Model

- Stars are point masses.
- Orbits are circular.
- Stars rotate synchronously.
- No radiation pressure effects.

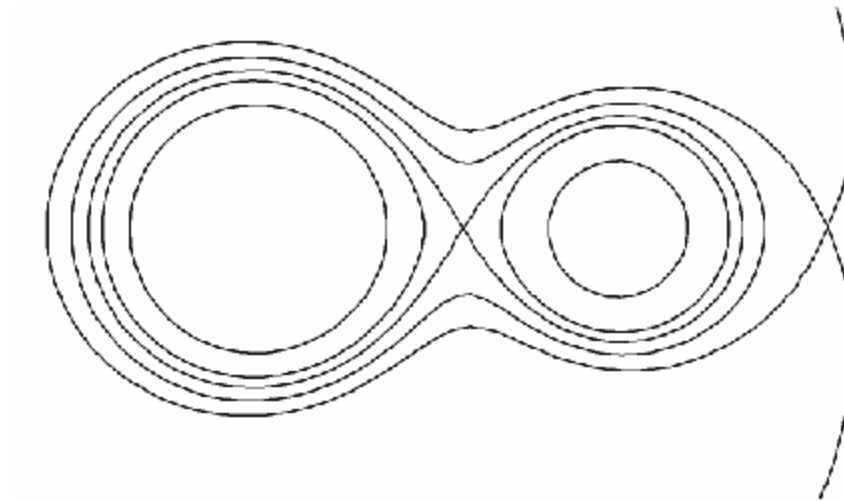


Tidal forces produce circular orbits and synchronous rotation in many interacting binaries.

Equipotentials are surfaces on which the sum of rotational and gravitational energy per unit mass is constant.

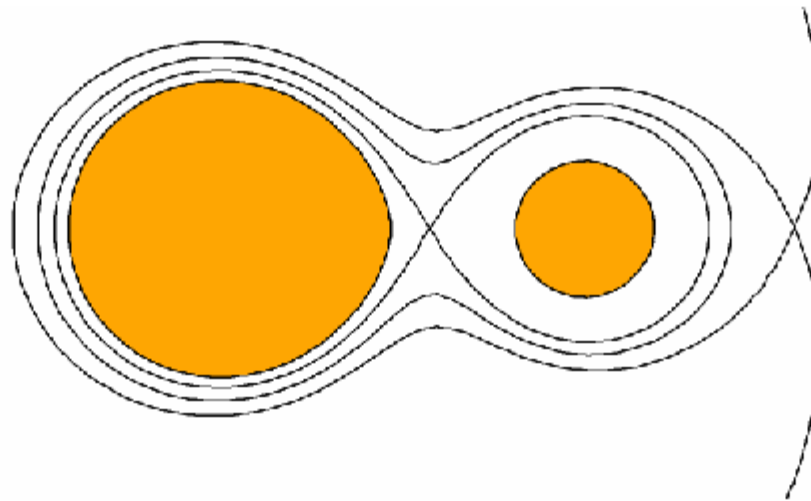
Belgrade, Jul 1, 2010

$$r = f(a, m_1/m_2)$$

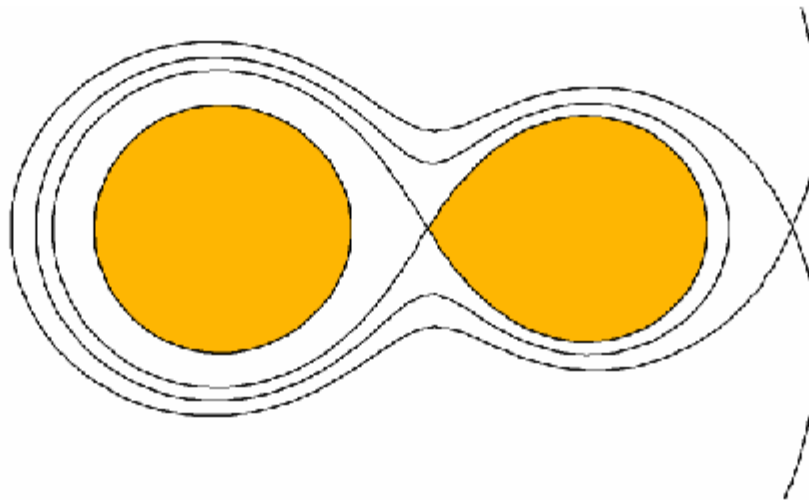


Belgrade, Jul 1, 2010

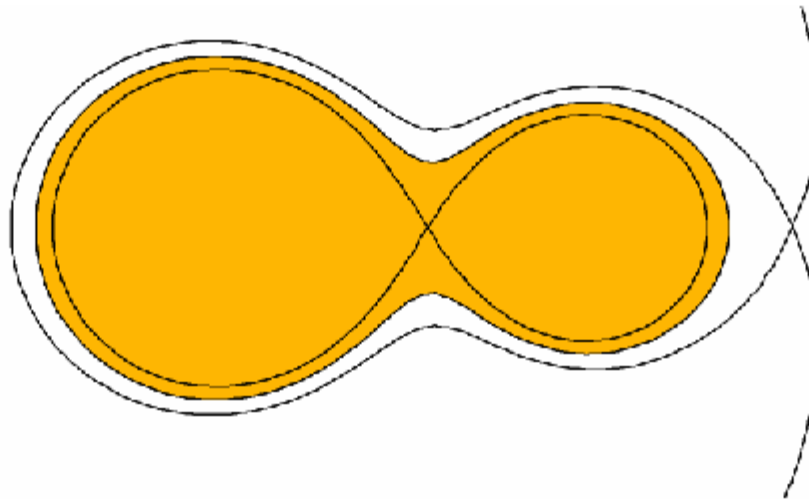
Detached (D): both components  
well within equipotentials



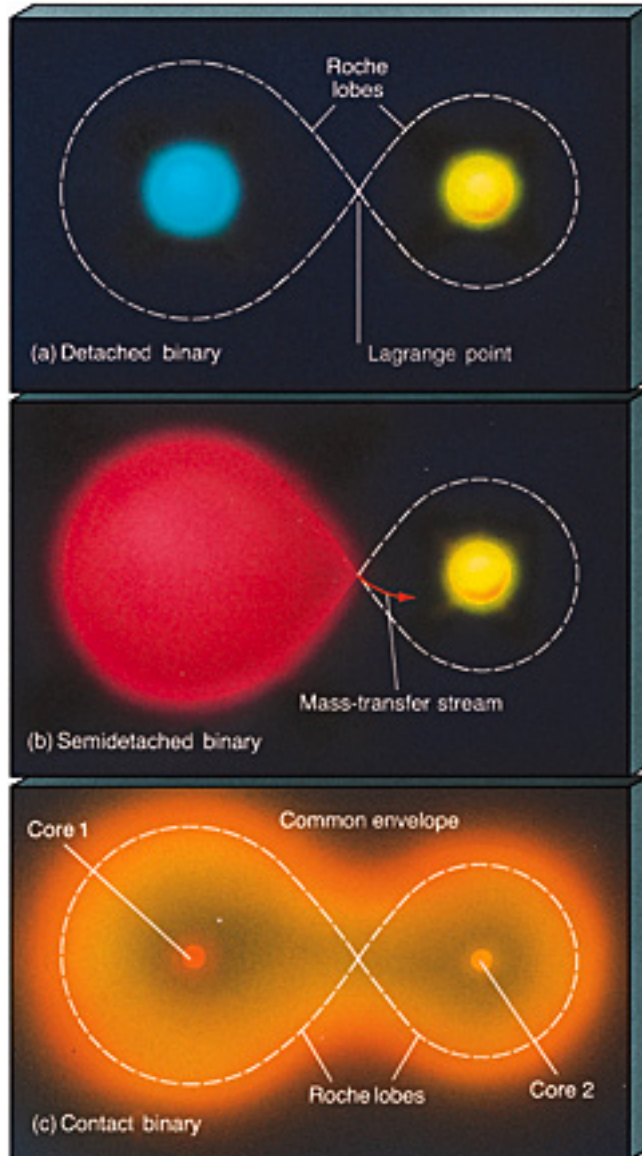
# Semi-detached (SD): one component reaches equipotential



Contact (C): both components  
reach equipotentials



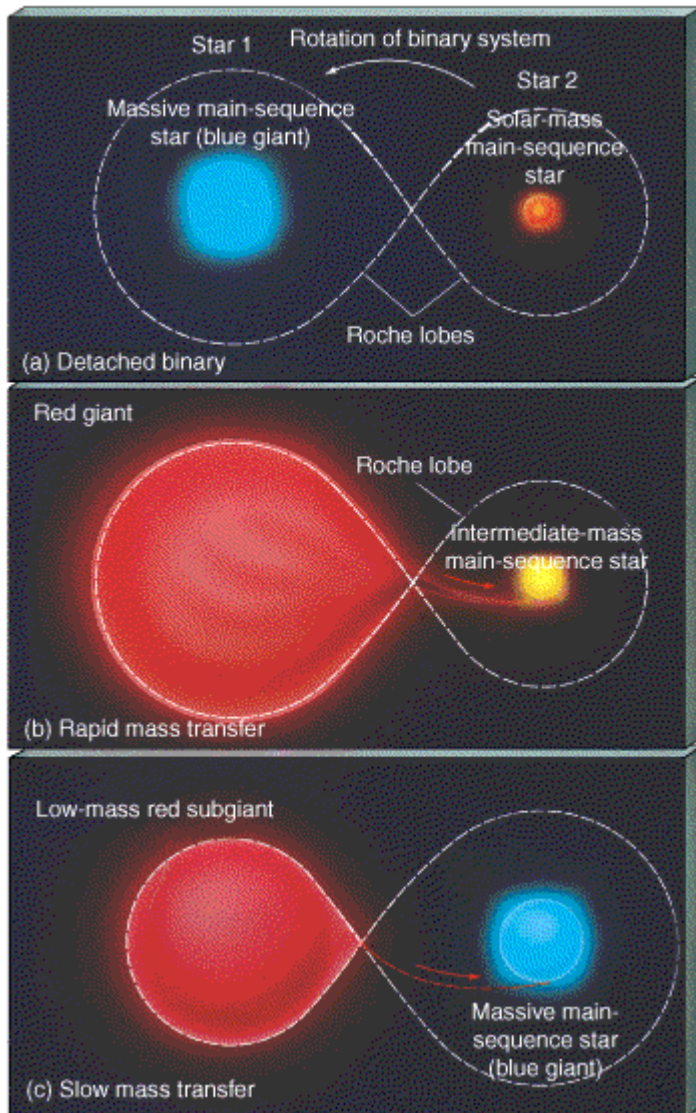
# Types: resume



- In a detached binary, each star lies within its respective Roche lobe.
- In a semidetached binary, one of the stars fills its Roche lobe and transfers matter onto the other, which still lies within its own Roche lobe.
- In a contact or common-envelope binary, both stars have overflowed their Roche lobes, and a single star with two distinct nuclear-burning cores results.

Belgrade, Jul 1, 2010

# The evolution of Algol



- Initially, Algol was probably a detached binary made up of two main-sequence stars —a relatively massive blue giant and a less massive companion similar to the Sun.
- As the more massive component (star 1) evolved off the main sequence it expanded to fill and eventually overflow its Roche lobe, transferring large amounts of matter onto its smaller companion (star 2).
- Today, star 2 is the more massive of the two, but it is on the main sequence. Star 1 is still in the subgiant phase and fills its Roche lobe, causing a steady stream of matter to pour onto its companion.

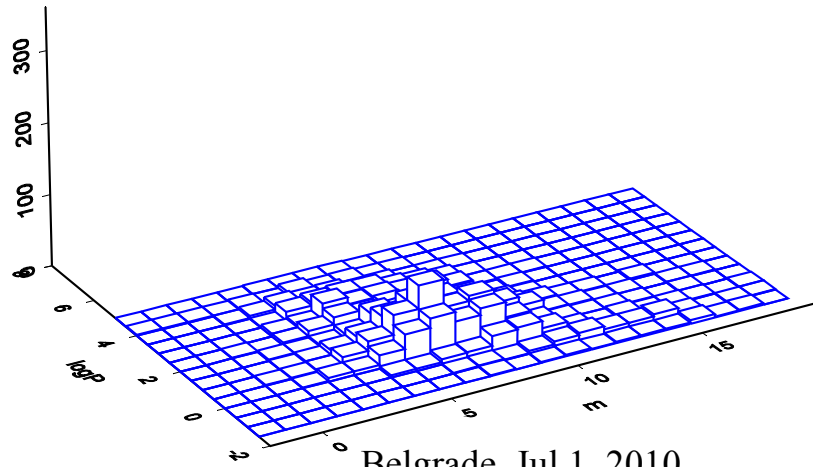
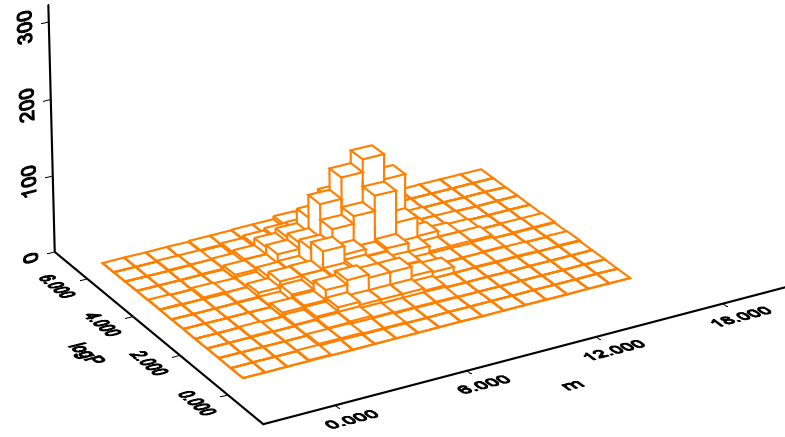
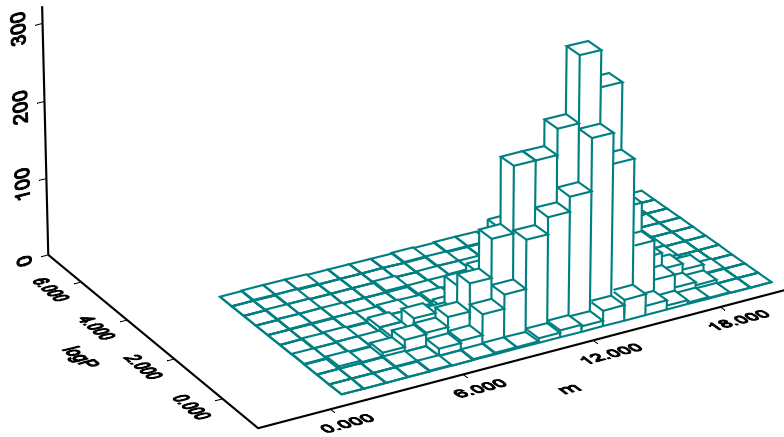
# Eclipsing binary stars

Belgrade, Jul 1, 2010

# Eclipsing binaries

- Eclipsing binaries are very numerous.
- Independent stellar mass and luminosity determination is possible only for components of eclipsing binaries with the spectrum lines of the two components. However, they represent only some 5% of all known eclipsing binaries.
- So, for statistical investigations it is necessary to estimate at least approximate values of fundamental parameters (such as mass and radius) for eclipsing binaries with unknown spectroscopic elements.

# Catalogued binaries: EB, VB (orbital), SB

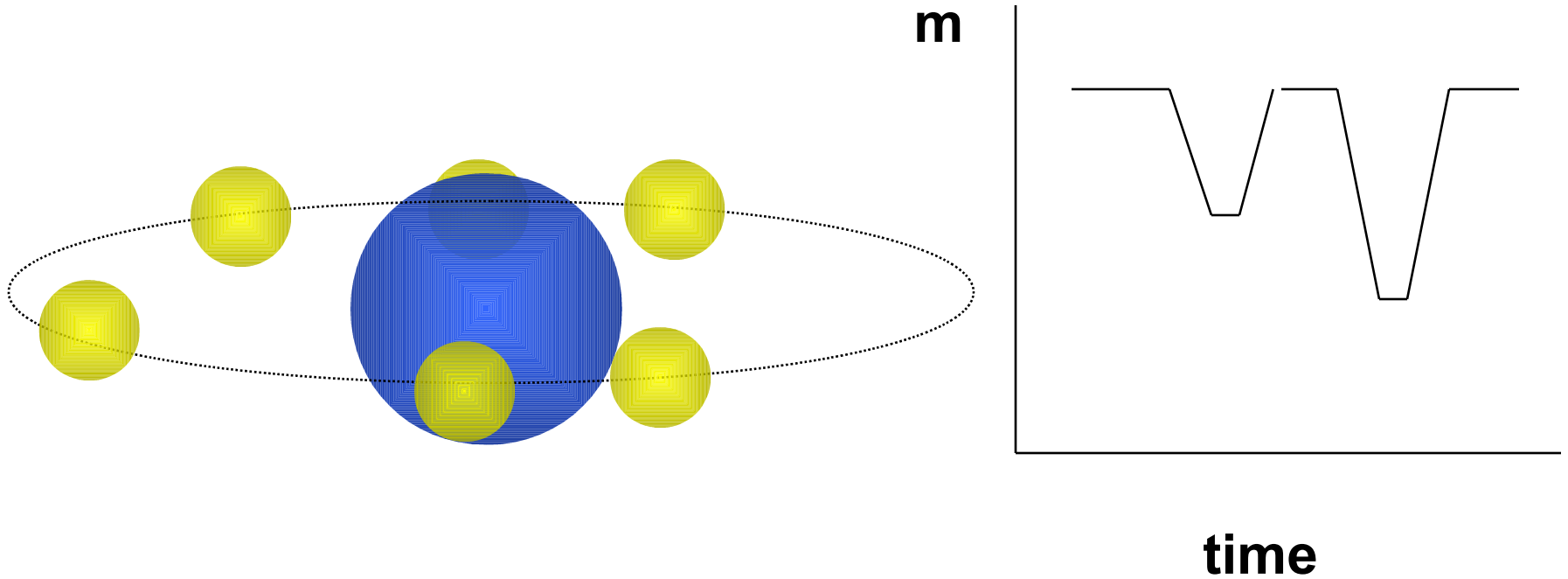


Belgrade, Jul 1, 2010

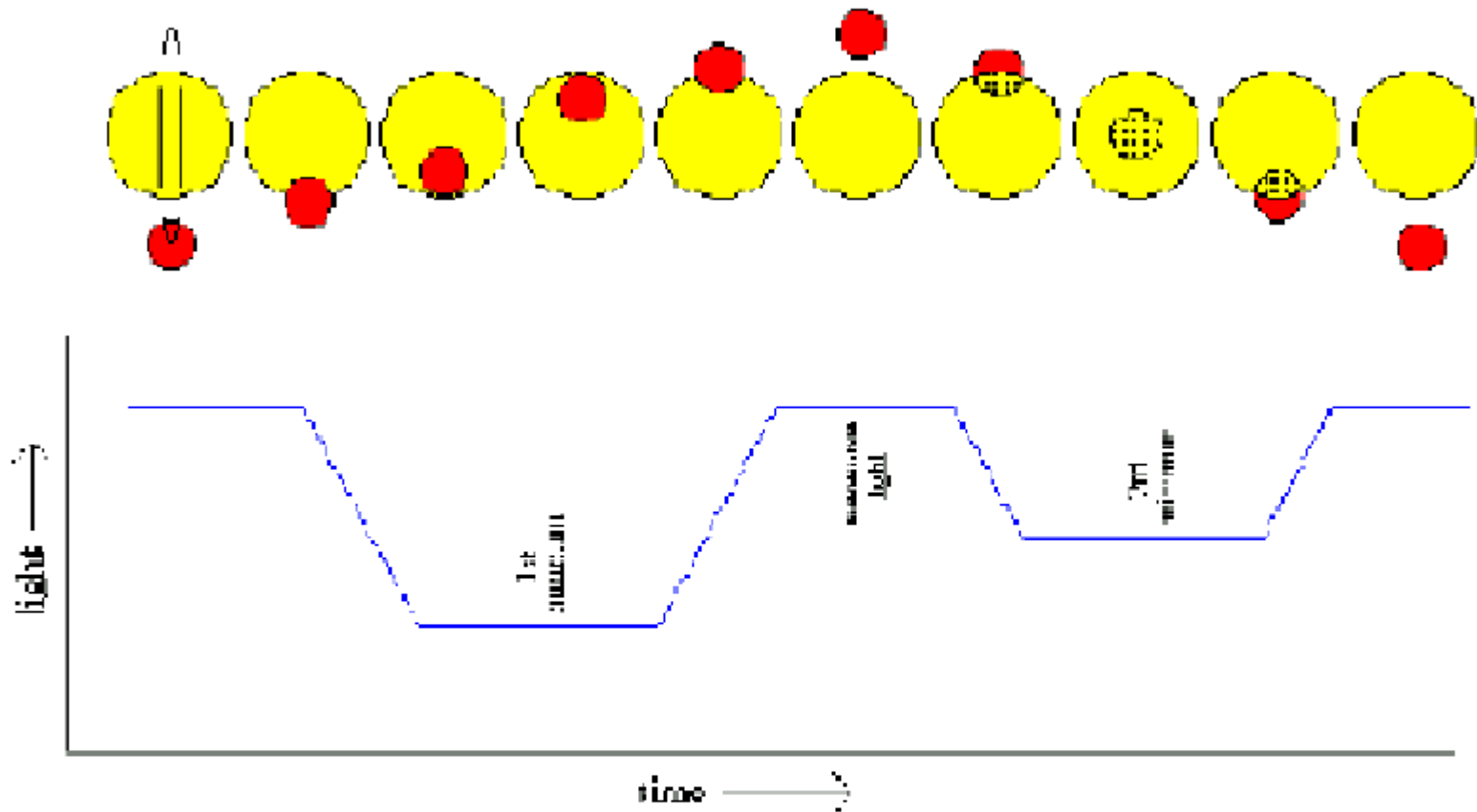
# Eclipsing binaries

- Eclipsing (photometric) binaries are binary stars of which one at times eclipses the other, thus leading to alterations in the apparent total brightness of the combined stars. The eclipse occurs because the line of sight lies almost in the orbital plane of the stars.
- The component stars can not be observed singly, the fact that they are binary stars being shown only by the variation caused by the eclipse, i.e., by photometric methods.

# Eclipsing binaries

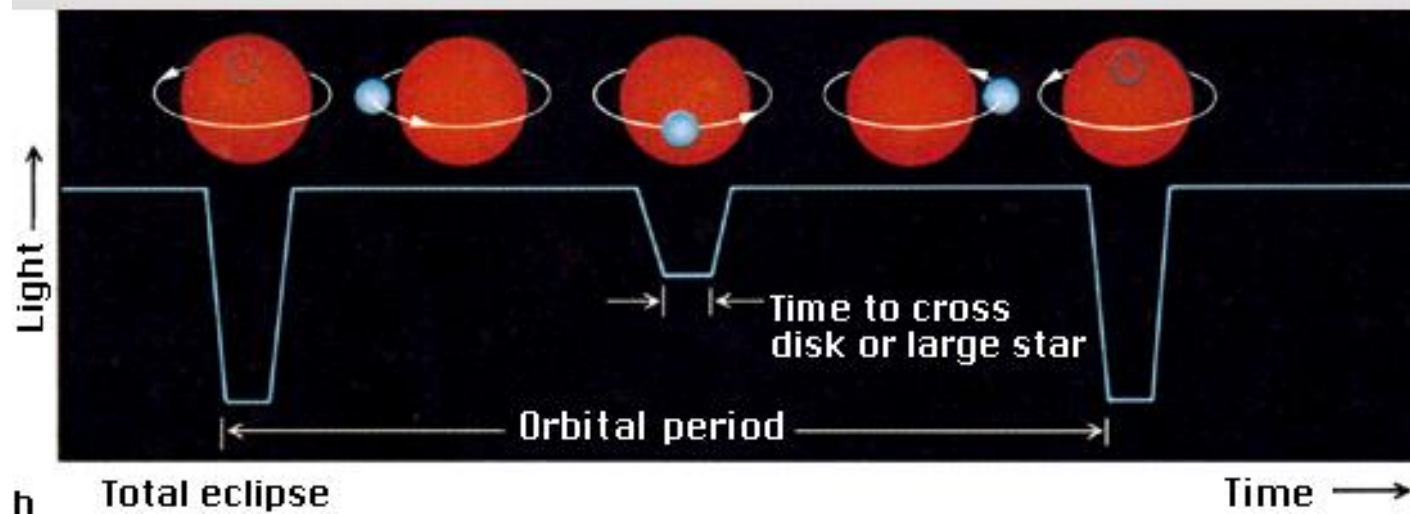
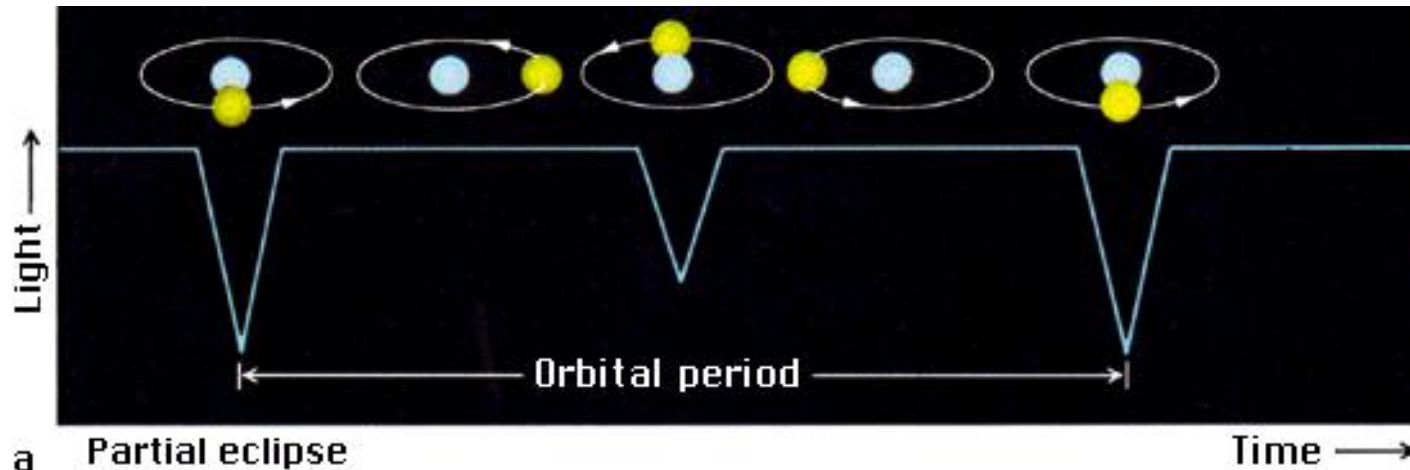


# Light curve



Belgrade, Jul 1, 2010

# Total and partial eclipses

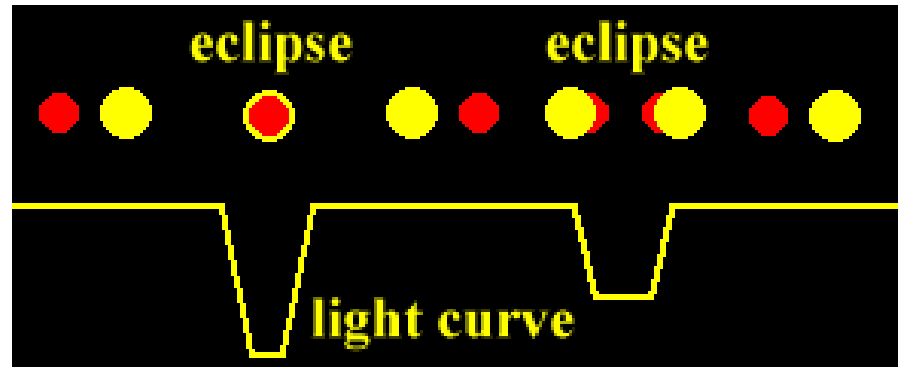


Belgrade, Jul 1, 2010

# Eclipses and transits

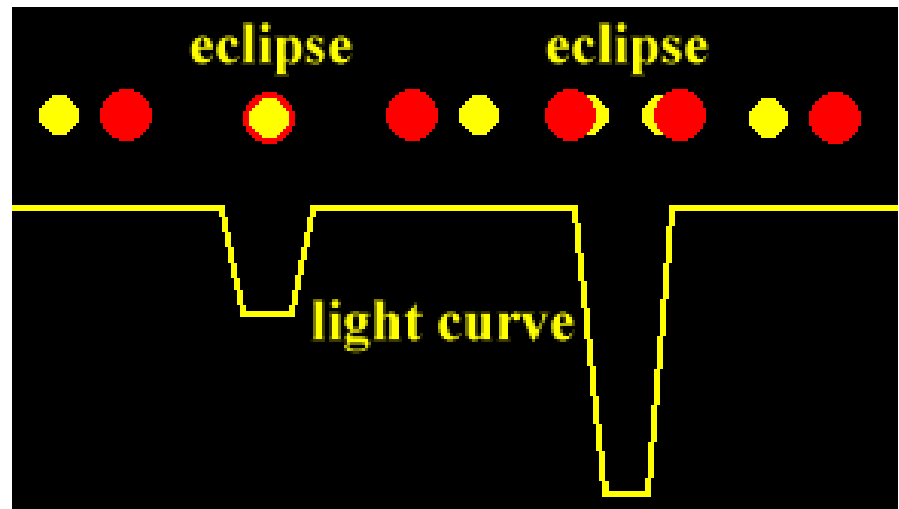
Larger component is  
the brighter one.

Primary minimum is  
the transit.

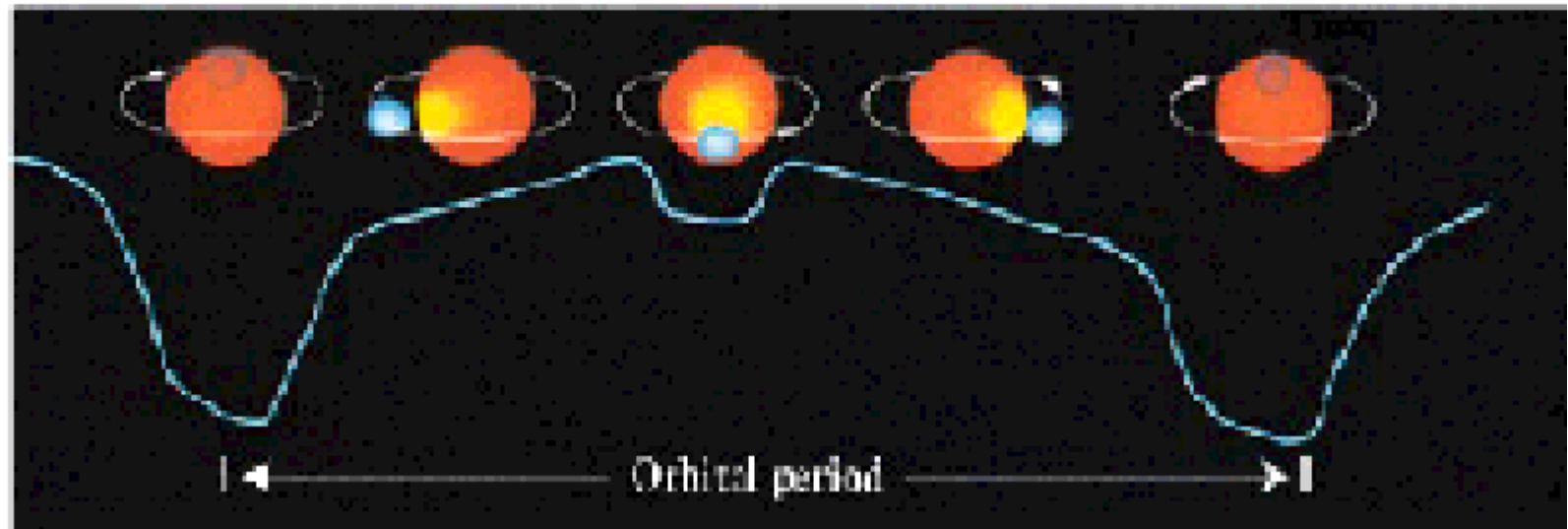
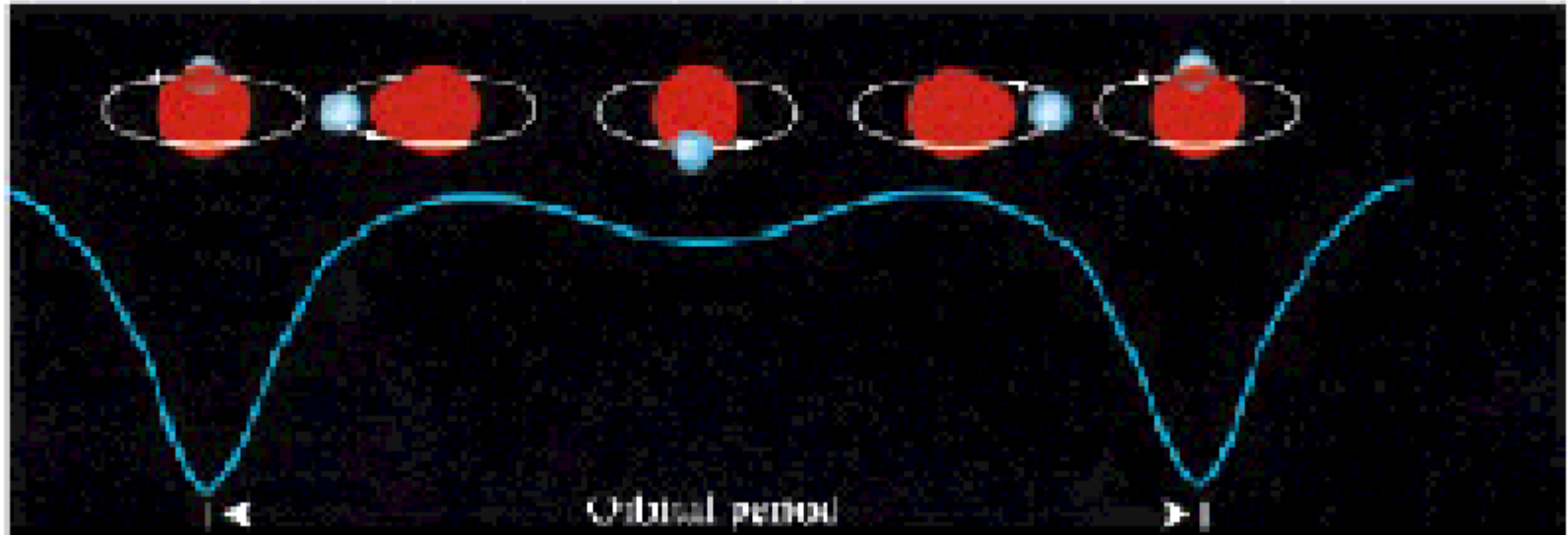


Larger component is  
the fainter one.

Primary minimum is  
the eclipse.

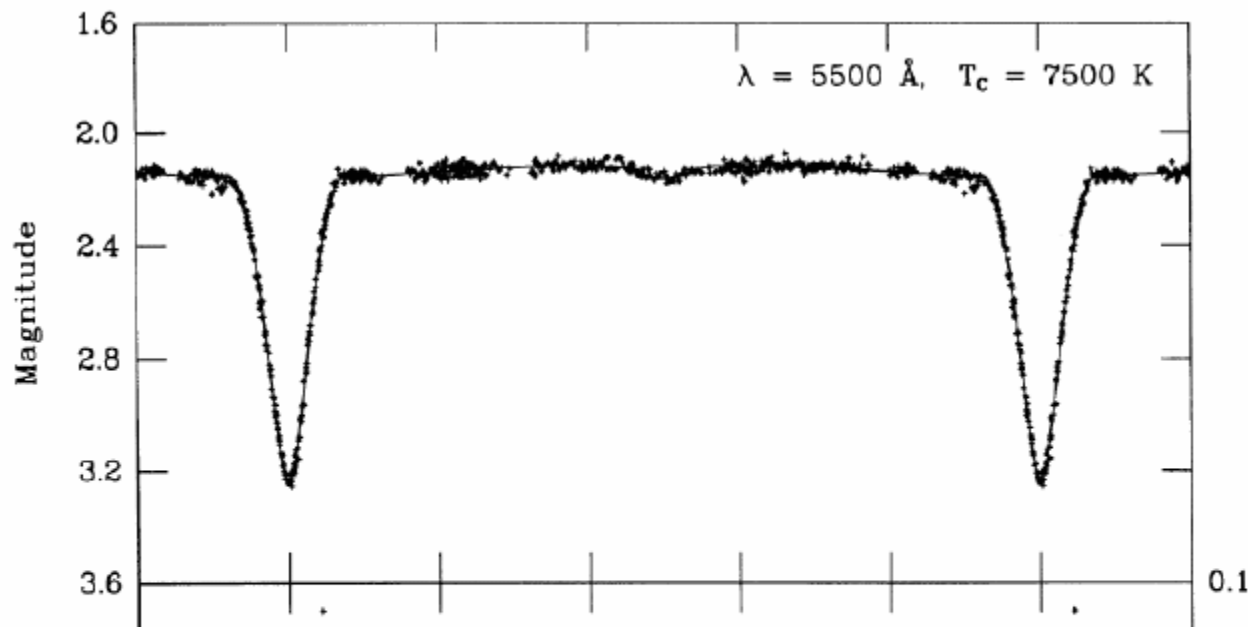


# Effects of tidal distortion and reflection

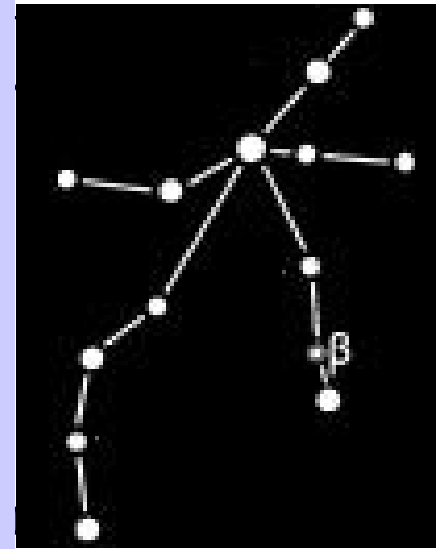


# Classification based on the shape of the light curve: EA, EB, EW

- simple
- traditional
- suits the observers



A  
1  
g  
0  
1

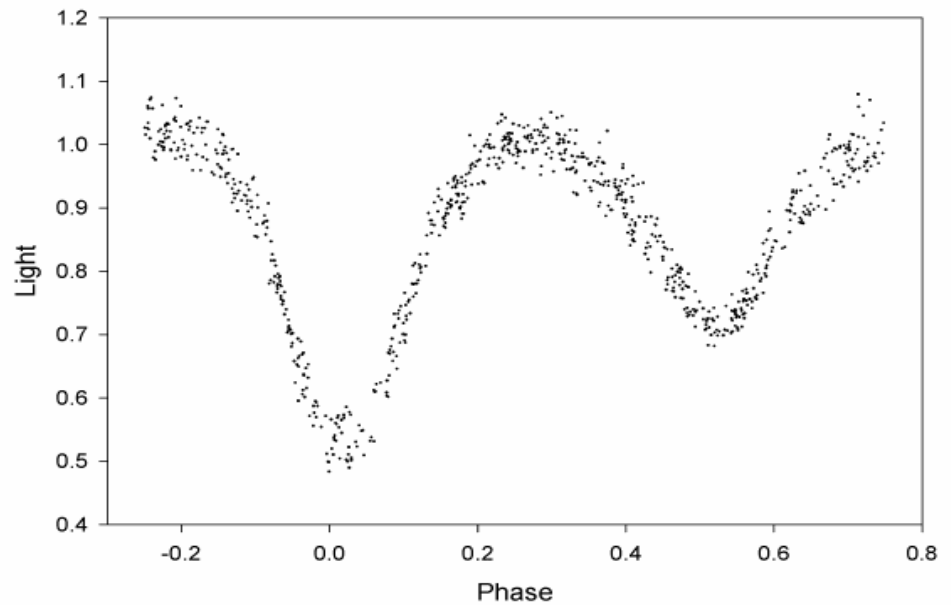


- It is possible to specify, for their light curves, the moments of the beginning and end of the eclipses.
- Between eclipses the light remains almost constant or varies insignificantly because of reflection effects, slight ellipsoidality of components, or physical variations.
- Secondary minima may be absent.
- An extremely wide range of periods is observed,  $0.2^d > P > 10000^d$ .
- Light amplitudes are also quite different and may reach several magnitudes.

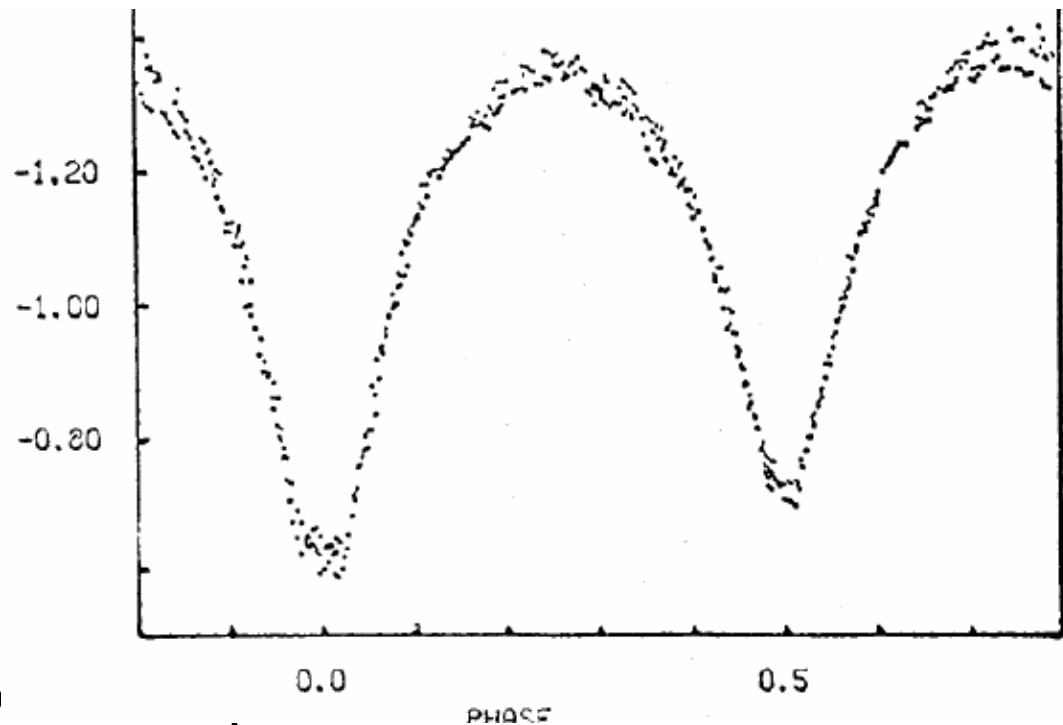
# EB: $\beta$ Lyrae

## – type eclipsing systems

- These are eclipsing systems
- It is impossible to specify, for their light curves, exact times of onset and end of eclipses because of a continuous change of a system's apparent combined brightness between eclipses
- Secondary minimum is observed in all cases, its depth usually being considerably smaller than that of the primary minimum
- Periods are mainly longer than 1 day.
- Light amplitudes are usually  $<2^m$  in V.
- The components generally belong to early spectral types (B-A).



# EW: W Ursae Majoris – type eclipsing systems



- Systems consist of ellipso
- Systems have light curves for which it is impossible to specify the exact times of onset and end of eclipses.
- The depths of the primary and secondary minima are almost equal or differ insignificantly.
- Light amplitudes are usually  $<0.^m8$ .
- Periods are shorter than 1 day
- The components generally belong to spectral types F-G and later.

# A light curve can provide relative quantities

- the relative radii of components (in units of  $a$ ):  $r_i = R_i/a$
- ratio of luminosities
- stellar figures
- center-to-limb variation of surface brightness (limb-darkening)
- third light (extra light of an optical or physical component)
- photometric mass ratio.

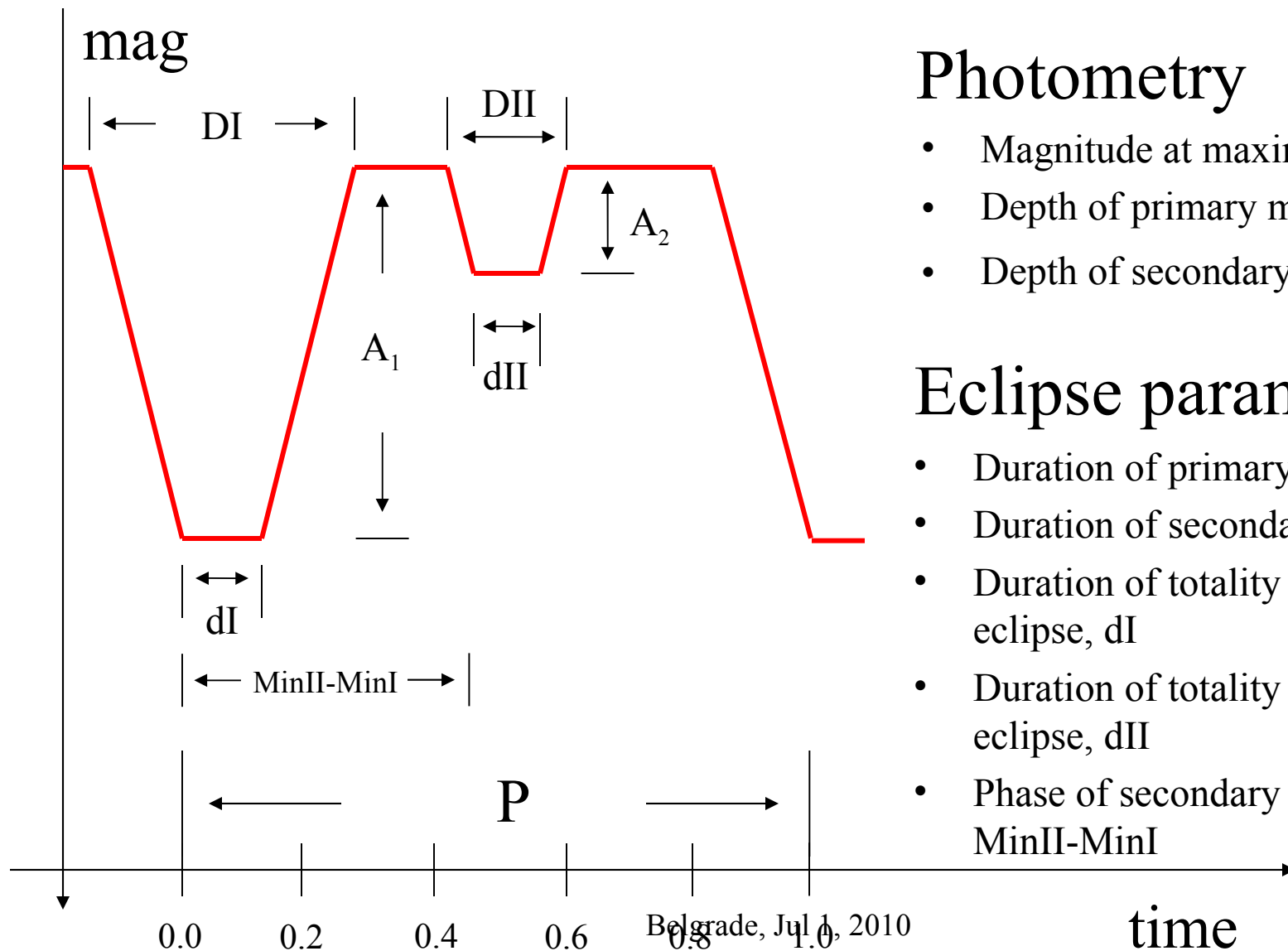
# Modern catalogues and lists of eclipsing binary stars

Belgrade, Jul 1, 2010

# Catalogues of eclipsing variables (light-curve elements)

- General Catalogue of Variable Stars (GCVS)
- A Finding List for Observers of Interacting Binary Systems, 5th Edition
- Eclipsing variables in microlensing surveys

# Light-curve elements



- Morphological type
- Epoch of primary minimum,  $T$
- Period,  $P$

## Photometry

- Magnitude at maximum brightness
- Depth of primary minimum,  $A_1$
- Depth of secondary minimum,  $A_2$

## Eclipse parameters

- Duration of primary eclipse,  $DI$
- Duration of secondary eclipse,  $DII$
- Duration of totality in primary eclipse,  $dI$
- Duration of totality in secondary eclipse,  $dII$
- Phase of secondary minimum,  $MinII-MinI$

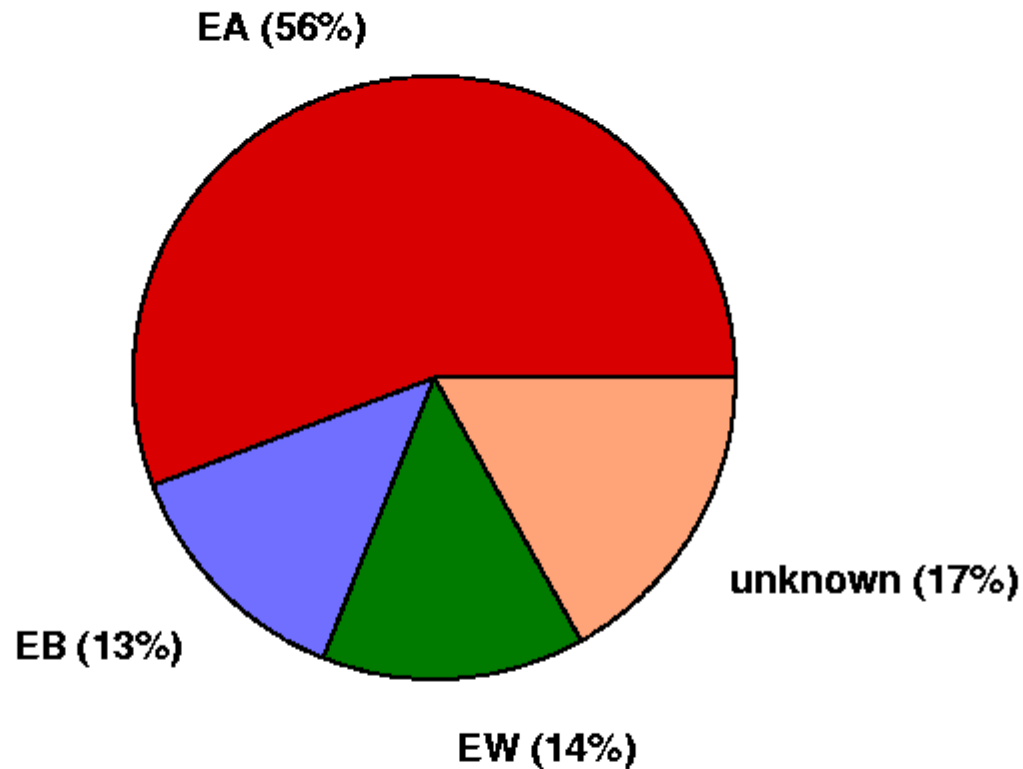
# General Catalogue of Variable Stars (GCVS)

- Kholopov P.N., Samus N.N., Frolov M.S., et al.
- GCVS contains data for more than 35,000 individual variable objects discovered and named as variable stars and located mainly in the Milky Way galaxy.
- In addition to star name and position, it contains the variable type, maximum and minimum magnitude, the epoch of maximum light, period, spectral type, and references.
- All variables in the data set are arranged in the order of their names inside constellations.

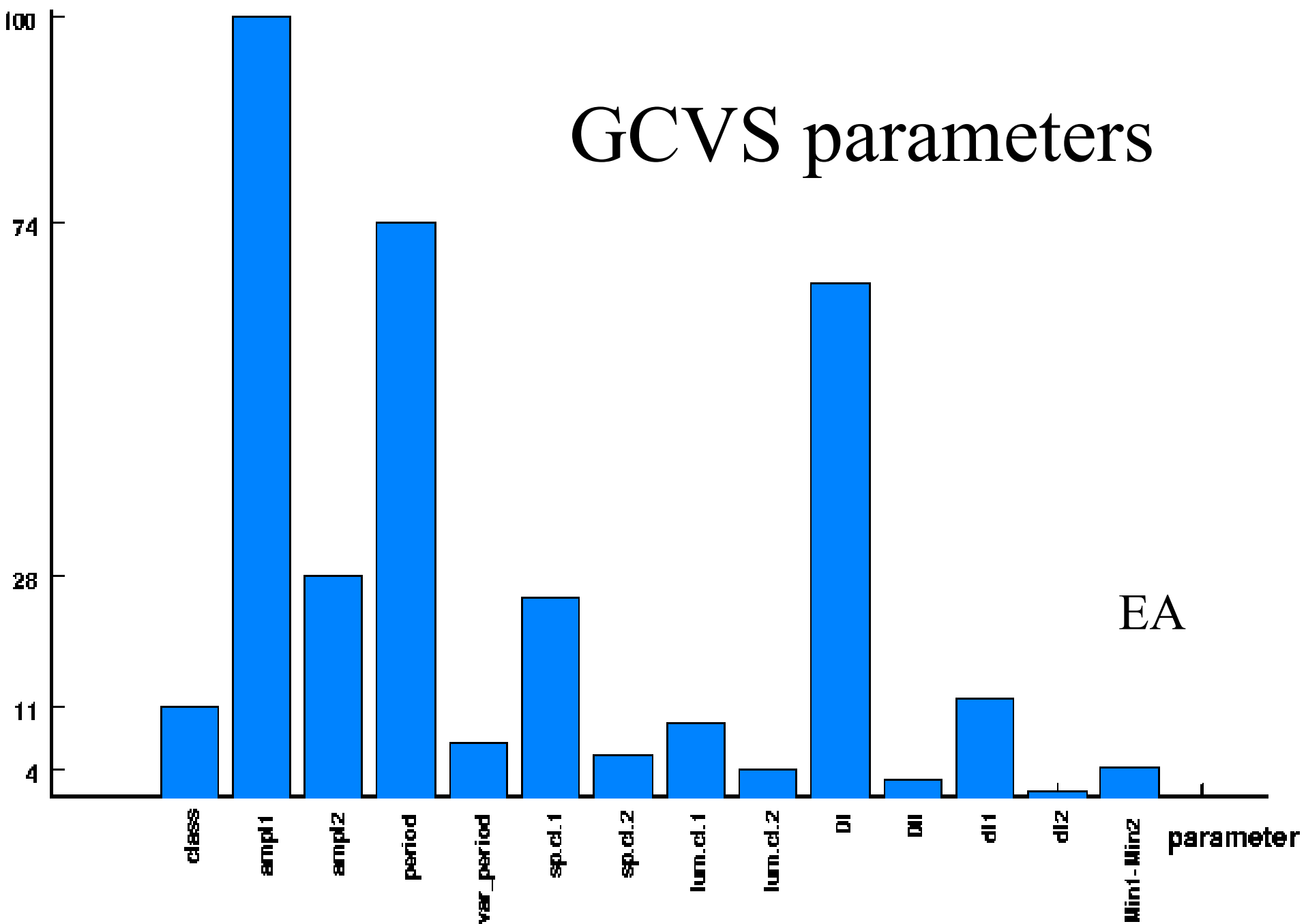
# GCVS contains

- an extensive series of cross-identification tables to alternative designations
- data for variables in external galaxies (including the Magellanic Clouds) and for extragalactic supernovae
- some 6300 eclipsing variables – current edition

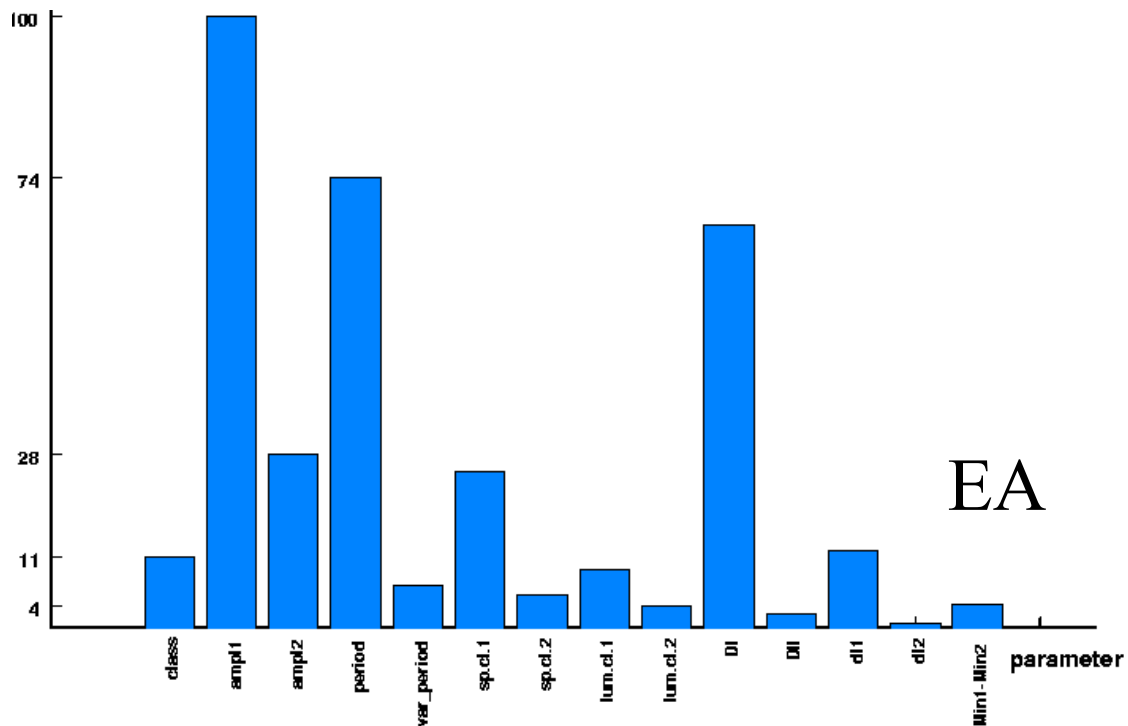
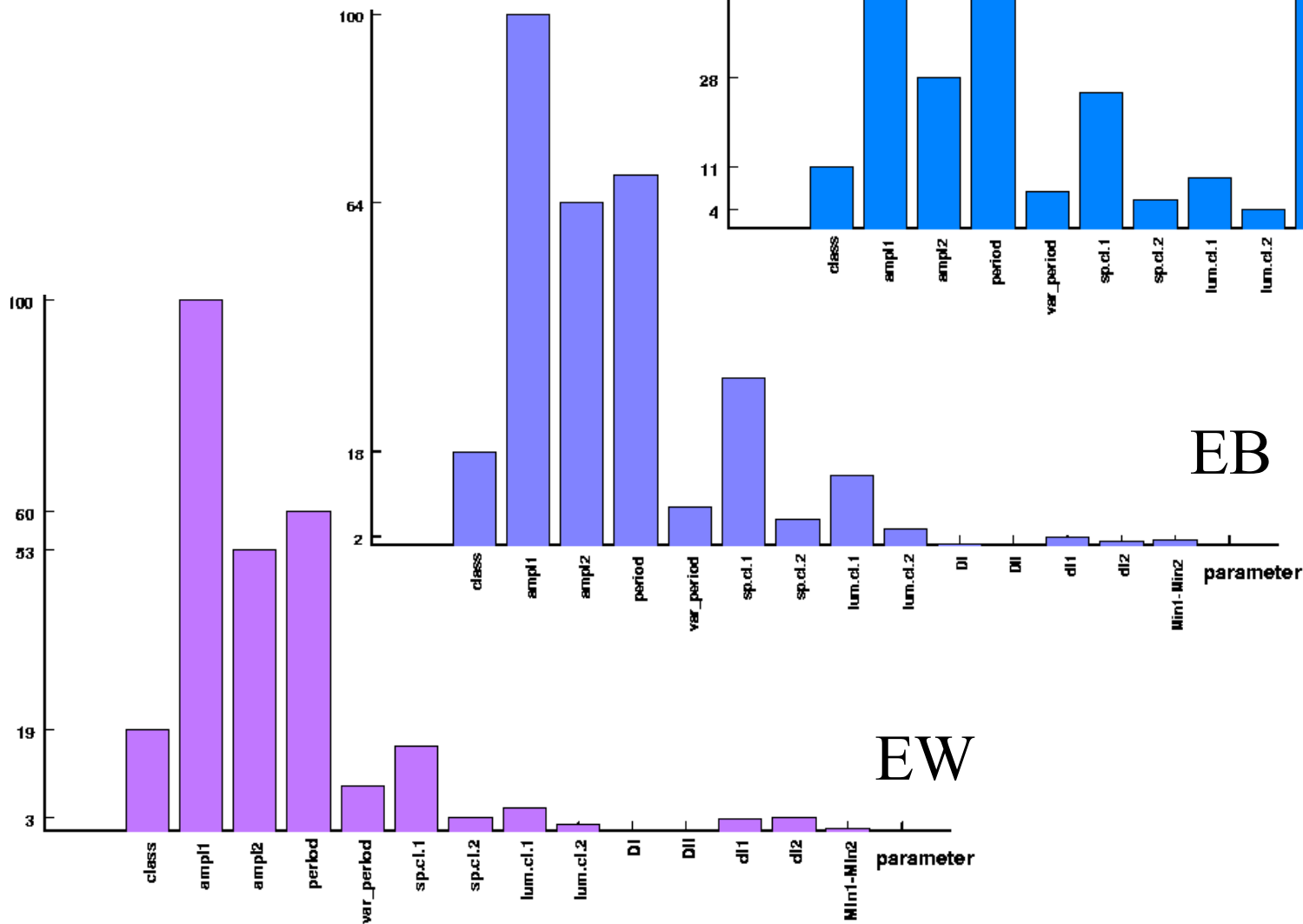
# GCVS stars: morphological type



# GCVS parameters



# GCVS parameters



EA

EB

EW

# A Finding List for Observers of Interacting Binary Systems, 5th Edition

- Wood F.B., Oliver J.P., Florkowski D.R., Koch R.H., 1980
- This catalogue is abstracted from the Card Catalog maintained at the University of Florida containing information on all published, and to the extent available, unpublished work on eclipsing binaries. The fifth edition differs from the previous ones in the extension of the magnitude limit at maximum light from 13<sup>m</sup> to 15<sup>m</sup>.
- The catalogue fields are: Finding List number; star name; position; magnitude at maximum light; depth of primary minimum; depth of secondary minimum; spectral class of star eclipsed at primary light; spectral class of star eclipsed at secondary light; epoch of primary minimum; orbital period; duration of primary minimum; duration of totality of primary minimum; BD, CoD, CPD, and HD number; alternate designations of system; codes indicating the nature of the system.
- The catalogue contains 3564 systems.

Belgrade, Jul 1, 2010

# Eclipsing variables in microlensing surveys

- The nature of microlensing searches is that millions of stars are monitored for the rare occasions when the light of one is amplified by an intermediate mass along the line-of-sight.
- Since about one star in a thousand is variable, the harvest of variable stars is far greater than the harvest of massive compact halo objects.
- Number of found variable stars is well over 100,000 – several times the number in the General Catalogue of Variable Stars

# Lists of eclipsing variables obtained as by-products of microlensing surveys

- The Optical Gravitational Lensing Experiment (OGLE)
- The MACHO Project (MAssive Compact Halo Objects)
- The All Sky Automated Survey (ASAS)
- Experience de Recherche d'Objets Sombres (EROS)
- Microlensing Observations in Astrophysics (MOA)
- Andromeda Galaxy Amplified Pixel Experiment (AGAPE)
- Disk Unseen Objects (DUO)
- Probing Lensing Anomalies NETwork (PLANET)

# Eclipsing variables statistics

- Number of known eclipsing variables:  
about 6300 in GCVS (3564 in Wood–5) and  
about 13000 in microlensing surveys (half  
of them are in LMC and SMC)

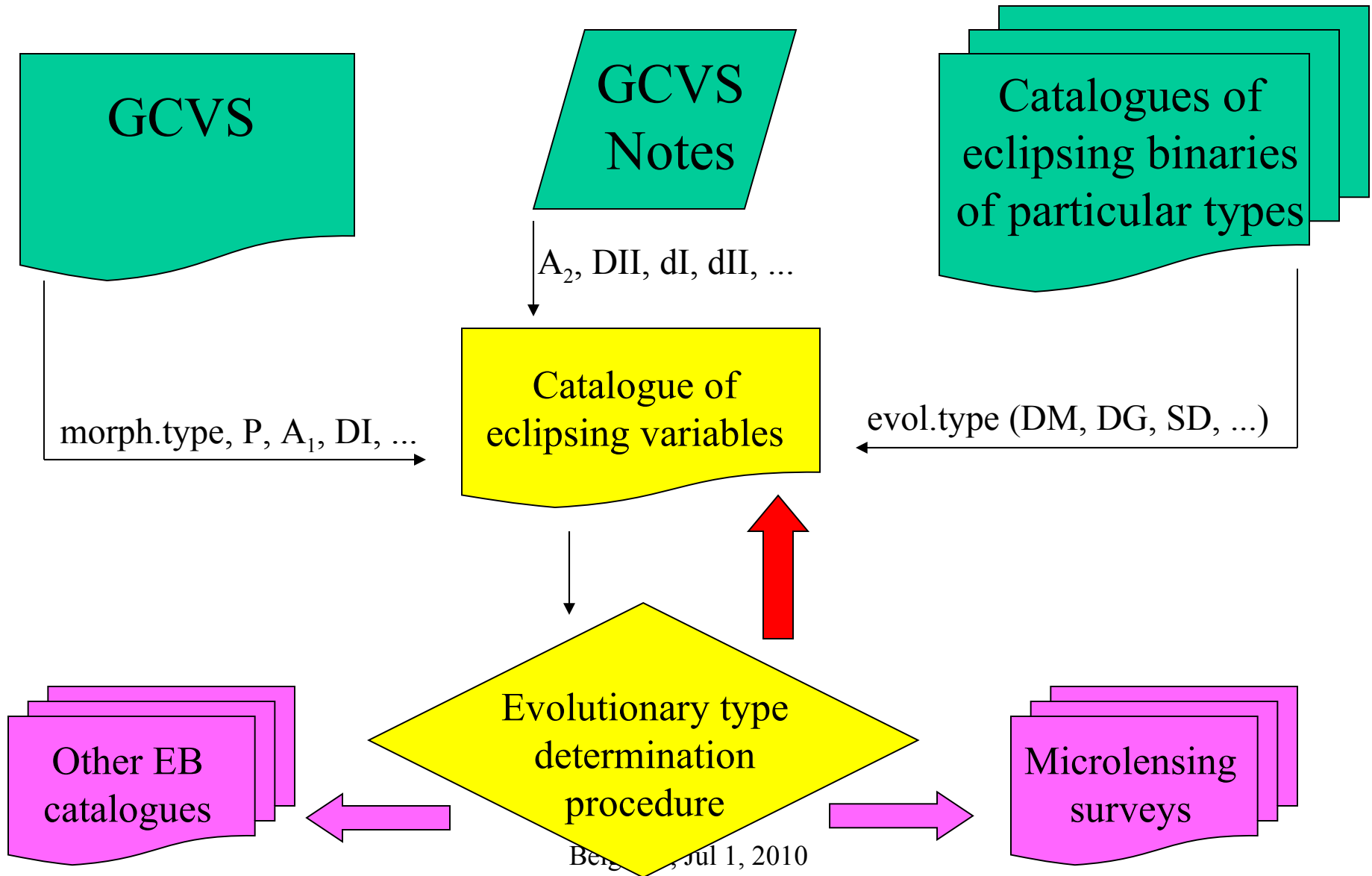
# Goals

- Determination of fundamental parameters (mass, radius, temperature, luminosity etc.) of eclipsing binaries.
- Construction of fundamental relations (e.g., mass-luminosity relation) and the initial mass function.
- Specification of evolutionary state of different system types and development of a self-consistent and comprehensive evolutionary scheme for close binaries.
- Study of marginal systems.

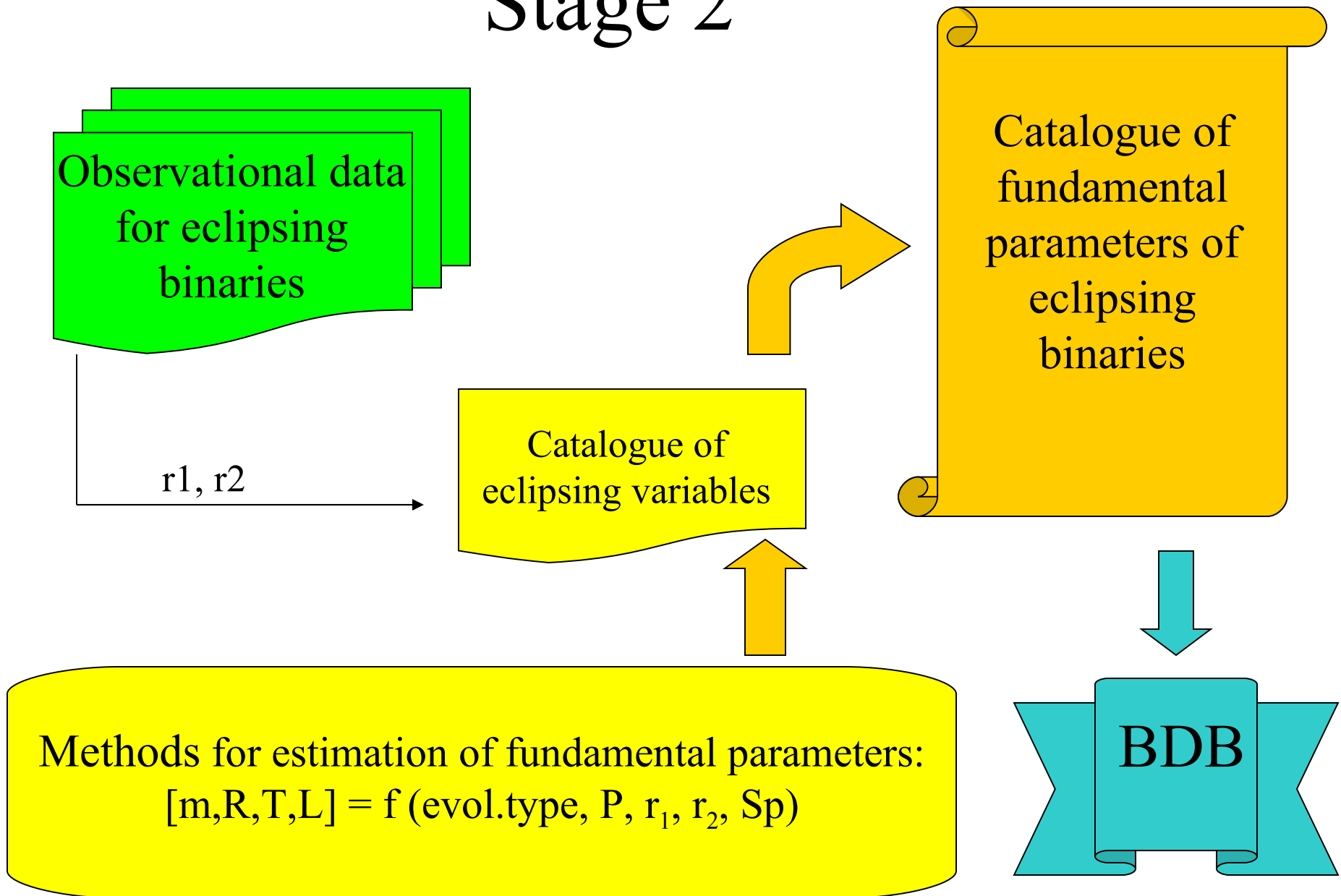
# How to do this?

- To compile a list of eclipsing variables, containing light-curve parameters and evolutionary type of systems.
- To construct a procedure for determination of system evolutionary type from observational data.
- To elaborate methods of determination of astrophysical parameters of eclipsing binary components from a limited set of observational data.
- To apply the methods to the list of eclipsing variables (with additional data on relative radii) and to construct a catalogue of fundamental parameters of eclipsing binaries.

# Stage 1



# Stage 2



# Extraction of relevant data from **GCVS** and its Notes

- morphological type of the light curve (EA, EB or EW)
- maximum and minimum magnitudes
- the photometric system for magnitudes
- period of the variable
- duration of primary eclipse, DI
- spectra of components

# Extraction of relevant data from GCVS and its Notes

- secondary minimum magnitude
- period variability indication
- duration of secondary eclipse, DII
- duration of totality in primary and in secondary eclipses, dI and dII
- phase of secondary eclipse, MinII-MinI

# Specification of a classification scheme for eclipsing binaries

- DM: Detached main sequence systems
- DR: Detached subgiant system
- DG: Detached giant or supergiant system
- S: Semi-detached system
- C: Contact system of unknown sub-class
- CB: Near-contact system of unknown sub-class
- CBF: Near-contact F system
- CBV: Near-contact V system
- CE: Early-type contact system
- CW: Late-type contact system of unknown sub-class
- CWA: Late-type contact A system
- CWW: Late-type contact W system

# Preparation for use of Svechnikov et al's catalogues

- The catalogue of orbital elements, masses and luminosities of **detached main-sequence** eclipsing variable stars with known photometric and spectroscopic elements. M.A.Svechnikov, E.L.Perevozkina, 1999.
- The catalogue of orbital elements, masses and luminosities of **detached main-sequence** eclipsing variable stars with known elements of photometric orbit and unknown spectroscopic elements. E.L.Perevozkina, M.A.Svechnikov, 1999.
- The catalogue of photometric, geometrical and absolute elements of **semidetached** eclipsing binary systems with known spectroscopic orbits. L.P.Surkova, M.A.Svechnikov, 2005.
- The catalogue of photometric, geometrical and absolute elements of **semidetached** eclipsing binary systems with known photometric orbits and unknown spectroscopic orbits. L.P.Surkova, M.A.Svechnikov, 2005.
- The catalogue of orbital elements, masses and luminosities of **short-period RS CVn** systems. G.N.Dryomova, E.L.Perevozkina, M.A.Svechnikov, 2004.

# Other catalogues of binaries – sources of independent classifiers

- Pribulla et al. (2003) – contact systems
- Budding et al. (2004) – semi-detached systems
- Shaw (1994) – near-contact systems
- Popper & Ulrich (1977) – subgiant detached systems
- Popper (1980), Malkov (1993) – various types

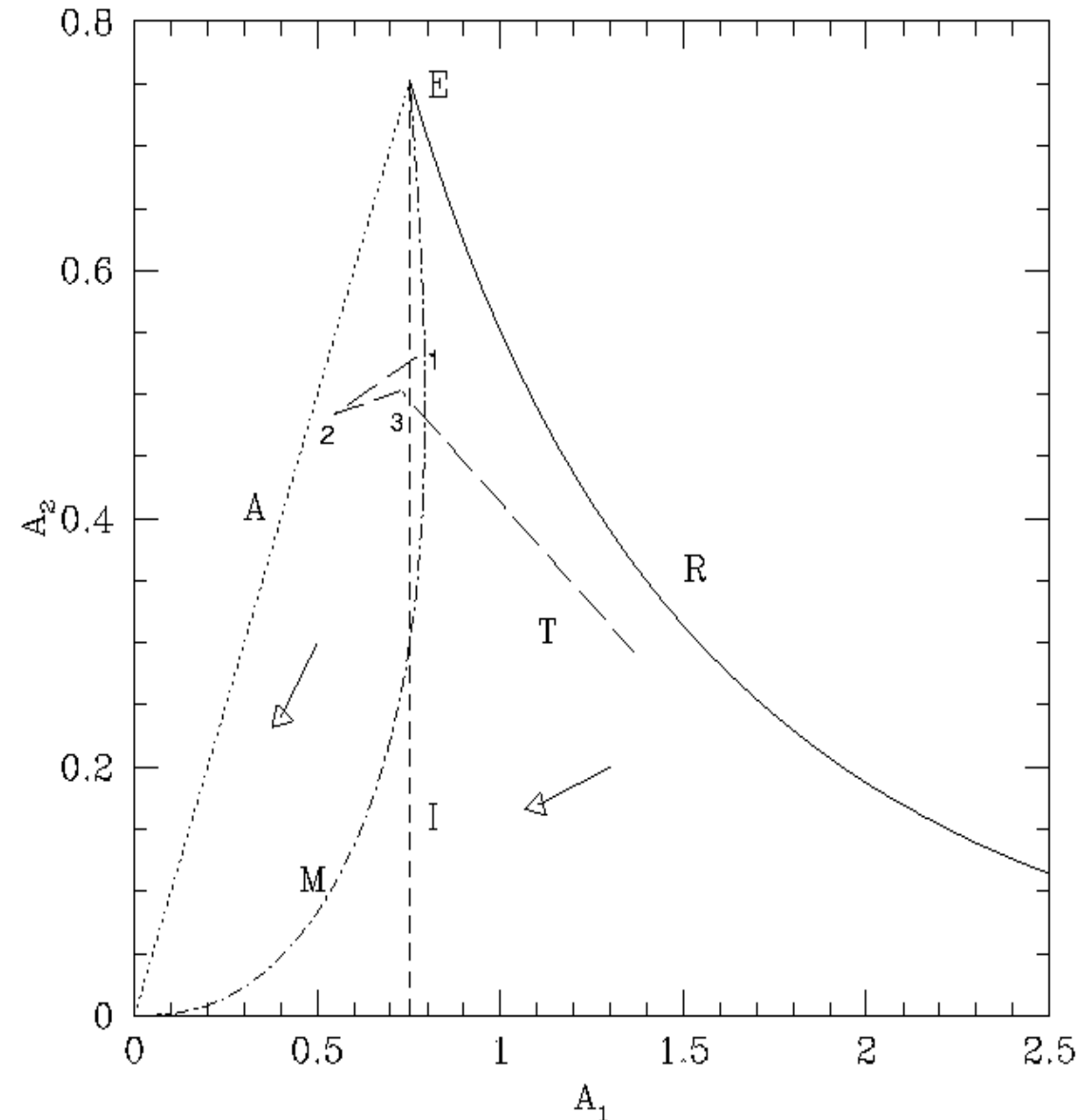
	A		B		C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	numb	name		class	mor	max	min1	A1	min2	A2	dA	ph	period	p	DI	dl	DII	dII	mm	ST1	LC1	ST2	LC2	
2	10012	RT	And	CB	EA	8,55	9,47	0,92	8,88	0,33	0,59	V	0,6289	d	170	0	170	0		480	5			
3	10025	SY	And	unknown	EA	10,7	12,2	1,5				V	34,9084	n	60	27				300		610		
4	10027	TT	And	S	EA	11,5	13	1,5	11,6	0,1	1,4	V	2,7651	n	140	0				300				
5	10030	TW	And	S	EA	8,8	10,86	2,06	8,94	0,14	1,92	V	4,1227	i	130	20				400	5	600		
6	10034	UU	And	S	EA	11,2	14,2	3				V	1,4862	v	220	0				450				
7	10045	WW	And	S	EA	10,3	11,4	1,1				V	23,2852	n	70	12				350		430		
8	10046	WX	And	unknown	EA	12,1	13,8	1,7				V	3,001	d	110	30								
9	10048	WZ	And	CB	EB	11,6	12,7	1,1	11,9	0,3	0,8	p	0,6956	d						450				
10	10051	XZ	And	S	EA	10,15	13,12	2,97	10,16	0,01	2,96	B	1,3572	i	210	0	260			340,34	4,5			
11	10055	AA	And	CB	EA	10,3	10,9	0,6	10,5	0,2	0,4	p	0,9351	n	210	0				280	5			
12	10056	AB	And	CWW	EW	9,5	10,32	0,82	10,2	0,7	0,12	V	0,3318	u						550		550	5	
13	10058	AD	And	unknown	EB	10,9	11,6	0,7	11,6	0,7	0	p	0,9861	i				494		300	5			
14	10066	AM	And	unknown	EA	12,5	13,7	1,2				p	8,8505	n	80									
15	10067	AN	And	DM	EB	6	6,16	0,16	6,09	0,09	0,07	p	3,2195	n						370	5			
16	10069	AP	And	DM	EA	11,3	11,9	0,6	11,9	0,6	0	p	1,5872	n	120	0				450				
17	10072	AS	And	unknown	EA	13,8	15,2	1,4				p	1,7001	n										
18	10082	BD	And	CW :	EB	11,3	11,7	0,4	11,4	0,1	0,3	p	0,4629	n						480				
19	10089	BL	And	CBV	EB	11	11,74	0,74	11,24	0,24	0,5	p	0,7223	d						300				
20	10092	BO	And	unknown	EA	13,4	16,3	2,9				p	5,7973	n	150					280				
21	10096	BS	And	unknown	EA	15	17	2				p		n										
22	10101	BX	And	CB	EW	8,9	9,57	0,67	9,15	0,25	0,42	p	0,6101	n						420	5			
23	10105	CD	And	unknown	EA	9,8	10,3	0,5				p	34,4416	n	120	0				480				
24	10114	CN	And	CB	EW	9,7	10,25	0,55	9,96	0,26	0,29	V	0,4627	n						480				
25	10115	CO	And	unknown	EA	11,1	12,1	1				p	1,8276	n	140	60				480				
26	10116	CP	And	unknown	EA	11,4	12,9	1,5	11,45	0,05	1,45	p	3,6089	n	130	10				350				
27	10121	CU	And	unknown	EA	12,5	16	3,5				p	1,7159	n	150	34								
28	10126	CZ	And	unknown	EA	12,4	13	0,6				p	2,7172	n	70					360				
29	10137	DO	And	unknown	E	12,2	13,1	0,9				p	0,6719	n										
30	10141	DS	And	CB	EB	10,4	10,9	0,5	10,7	0,3	0,2	V	1,0105	n						420	3			
31	10145	DW	And	unknown	E	13,6	14,4	0,8				p		n										
32	10155	EL	And	unknown	E	12,7	13,5	0,8				p		n										
33	10159	EP	And	CWW	EW	11,9	12,5	0,6	12,5	0,6	0	p	0,4041	n										

# Construction of a new procedure: list of rules used for classification

- Light curve morphological type allowable values
- $A_1$ - $A_2$  (depths of primary and secondary minima) relation for detached systems
- $A_1$ - $A_2$  relation for detached MS systems
- $A_1$ . $A_2$  and  $dA (=A_1-A_2)$  upper limits
- $P$  (period) upper and lower limits
- $P$  variation allowable values
- Duration of eclipses difference (DI-DII) ranges
- Phase of secondary minimum MinII-MinI ranges
- Primary and secondary spectral type ( $Sp_1$ ,  $Sp_2$ ) ranges
- Primary and secondary luminosity class ( $LC_1$ ,  $LC_2$ ) ranges

Belgrade, Jul 1, 2010

# Theoretical location of detached systems with total eclipses in the $A_1$ - $A_2$ plane



**A**: systems with  $A_1 = A_2$ .

**R**: systems with components of equal radii, upper limit for detached systems.

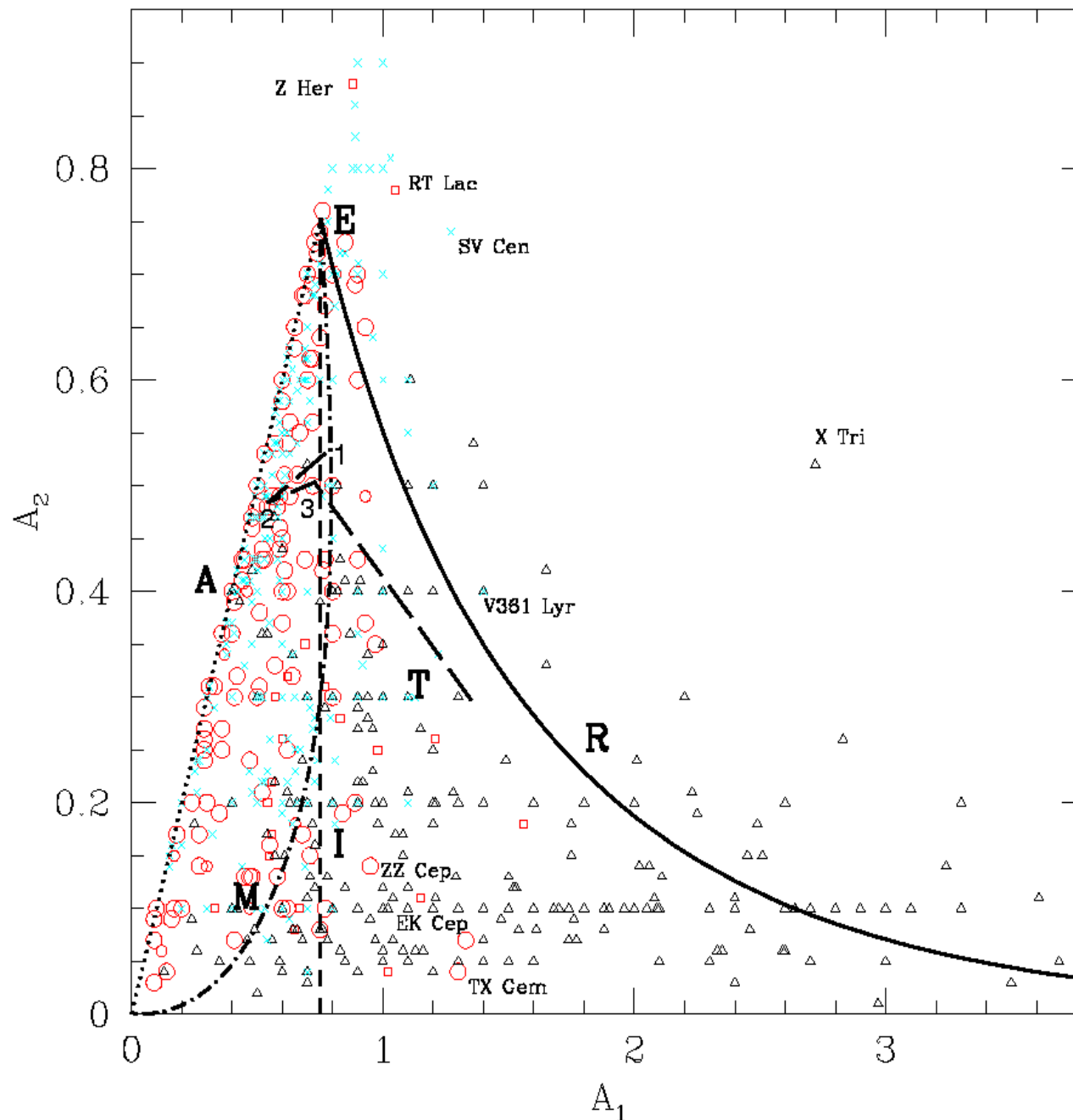
**I**: left limit for systems, where brighter component is smaller.

**M**: approximate relation for MS stars.

**T**: evolutionary track of a  $2.8m_{\odot} + 2.5m_{\odot}$  system.

**E**: systems with identical components.

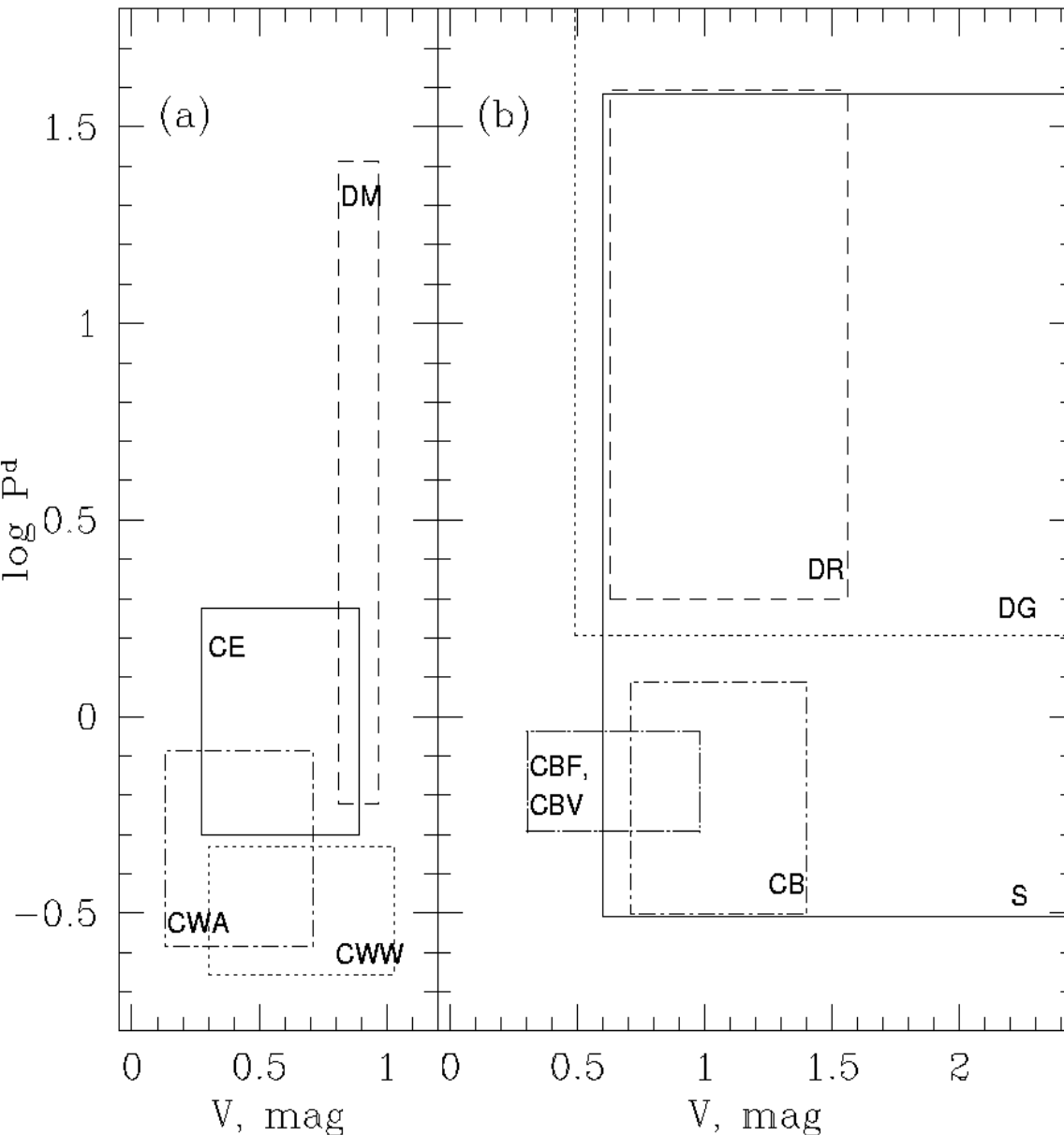
Arrows indicate direction of a system movement on the diagram if the eclipse becomes partial.



Catalogued  
systems in the  
 $A_1$ - $A_2$  plane

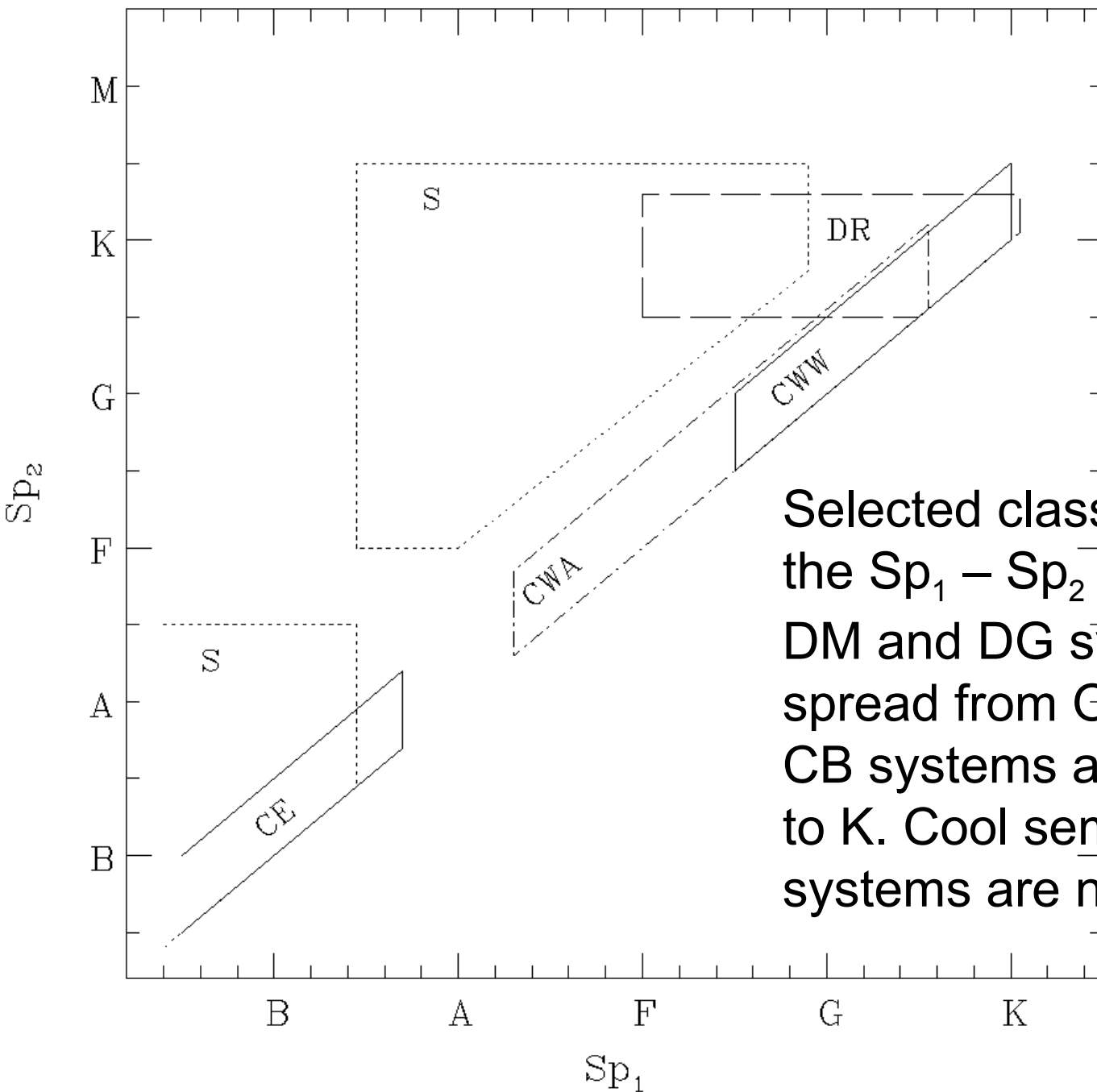
circles --- DM  
systems  
(larger), DR  
systems  
(smaller),  
squares --- DR  
systems,  
triangles --- S  
systems,  
crosses --- C  
systems.

# Limiting amplitudes and periods



Lower and upper borders of rectangles indicate minimum and maximum period for a given class, respectively. Right border of rectangles indicates maximum value for the depth of primary minimum,  $A_1$ .

Left border of rectangles indicates maximum value for the depths difference,  $dA$  (for DM, CE and CW systems, panel (a) or maximum value for the depth of secondary minimum,  $A_2$  (for DR, DG, S and CB systems, panel (b)).



Selected classes of systems in the  $Sp_1 - Sp_2$  plane. Spectra of DM and DG systems are spread from O to M. Spectra of CB systems are spread from B to K. Cool semi-detached systems are not shown.

# How the procedure works

- The procedure assigns an evolutionary class to a system, basing on its available observational parameters. One of the following classes can be assigned: D (detached system of unknown sub-class), DM, DR, DG, S, C, CB, CBV, CBF, CE, CW, CWA, CWW.
- If, according to the result of the classification, a system can belong to more than one of the listed classes, its class is considered to remain unknown. However, if a system could belong to both DM and DR, it was classified as D, etc.

# Verification of the procedure, using 1029 stars with known classifier

- Altogether 475 systems (46%) were classified, class of others remains unknown.
- In 189 cases a less accurate classifier was assigned to a system (e.g., CWW  $\rightarrow$  CW).
- For 19 systems a classifier was made more accurate (e.g., CB  $\rightarrow$  CBV).
- In general
  - 113 of 194 (58%) D systems,
  - 79 of 437 (18%) S systems and
  - 283 of 398 (71%) C systemswere classified correctly.

# Classification of 5301 eclipsing variables with unknown classifier

- For 86 stars all classes are forbidden (remain unclassified)
- More than one class can be applied to 4225 (class remains unknown)
- 990 stars are successfully classified

# Reasons for the failure of the classification process

- Star belongs to a marginal class: X-ray binary (V1341 Cyg), cataclysmic variable (RW Tri), polar (EF Eri), symbiotic star (V1329 Cyg)
- Star has constant light (NW Aur)
- Star has a period that is twice (or half) as much as catalogued one (RU Ind, AP Aps, HV TrA)
- Other errors in the GCVS

# Statistics of 990 successfully classified stars

- 188 detached binaries, among them:
  - 58 detached MS systems
  - 1 detached subgiant system
  - 81 detached giant or supergiant systems
- 199 semidetached binaries
- 603 contact and near-contact binaries, among them:
  - 36 near-contact systems (3 of them are CBF)
  - 7 early type contact systems
  - 39 late type contact systems (24 of them are CWW)

# Classification of EB from large catalogues

- Catalogue of eclipsing variables (Malkov et al. 2006): light curve morphological type,  $A_1$ ,  $A_2$ ,  $Sp_1$ ,  $Sp_2$ ,  $LC_1$ ,  $LC_2$ , P, P variation information, DI, DII, MinII-MinI for 5301 stars
- Wood et al. (1980):  $A_1$ ,  $A_2$ ,  $Sp_1$ ,  $Sp_2$ ,  $LC_1$ ,  $LC_2$ , P, DI, DII for 3564 stars

# Classification of EB from large microlensing surveys

- OGLE-LMS and OGLE-SMC: light curve morphological type,  $A_1$ ,  $A_2$ ,  $P$ , MinII-MinI for 4085 stars
- MACHO:  $A_1$ ,  $P$ , MinII-MinI for 6143 stars
- ASAS-3:  $A_1$ ,  $P$  for 15481 stars

Large (more than thousand EB systems) surveys with at least two parameters are chosen

# Results of classification.

<b>Survey</b>	<b>No of stars</b>	<b>No of parameters</b>	<b>D</b>	<b>S</b>	<b>C</b>
<b>CEV</b>	5301	13	188	199	603
<b>Wood-5</b>	3564	10	193	166	101
<b>OGLE-LMC</b>	2681	6	559	69	78
<b>OGLE-SMC</b>	1404	6	388	16	135
<b>MACHO</b>	6143	3	1125	24	25
<b>ASAS-3</b>	15481	2	536	73	811

For the majority of the classified stars  
more detailed classification is available

# Current activity

- Improvement of the classification scheme
- Correction of the catalogue data
- Specification of rules for exotic (marginal) systems
- Use of other algorithms for classification: Bayes classifiers, decision trees etc.; in collaboration with Institute of Informatics Problems (IIP) RAS and Institute for Information Transmission (IITP) Problems RAS
- Development of methods of parameterisation
- Compilation of relevant catalogues

# Data Mining as a Part of RVO Infrastructure

- Technical part of the work on eclipsing binaries classification is considered in the frame of the Russian Virtual Observatory.
- Well organised part of the RVO infrastructure is based on the AstroGrid system developed in UK and installed in Moscow at the Supercomputer Centre of the Russian Academy of Sciences and at the Institute of Informatics Problems of the Russian Academy of Sciences. These two installations are intended for problem solving by the Russian astronomers.
- RVO infrastructure is gradually extended with new instruments that should provide for solving of various problems of data analysis. In this connection an issue of incorporation into the RVO infrastructure of Data Mining tools is important.
- Eclipsing binaries classification problem played a catalyst role in this process.

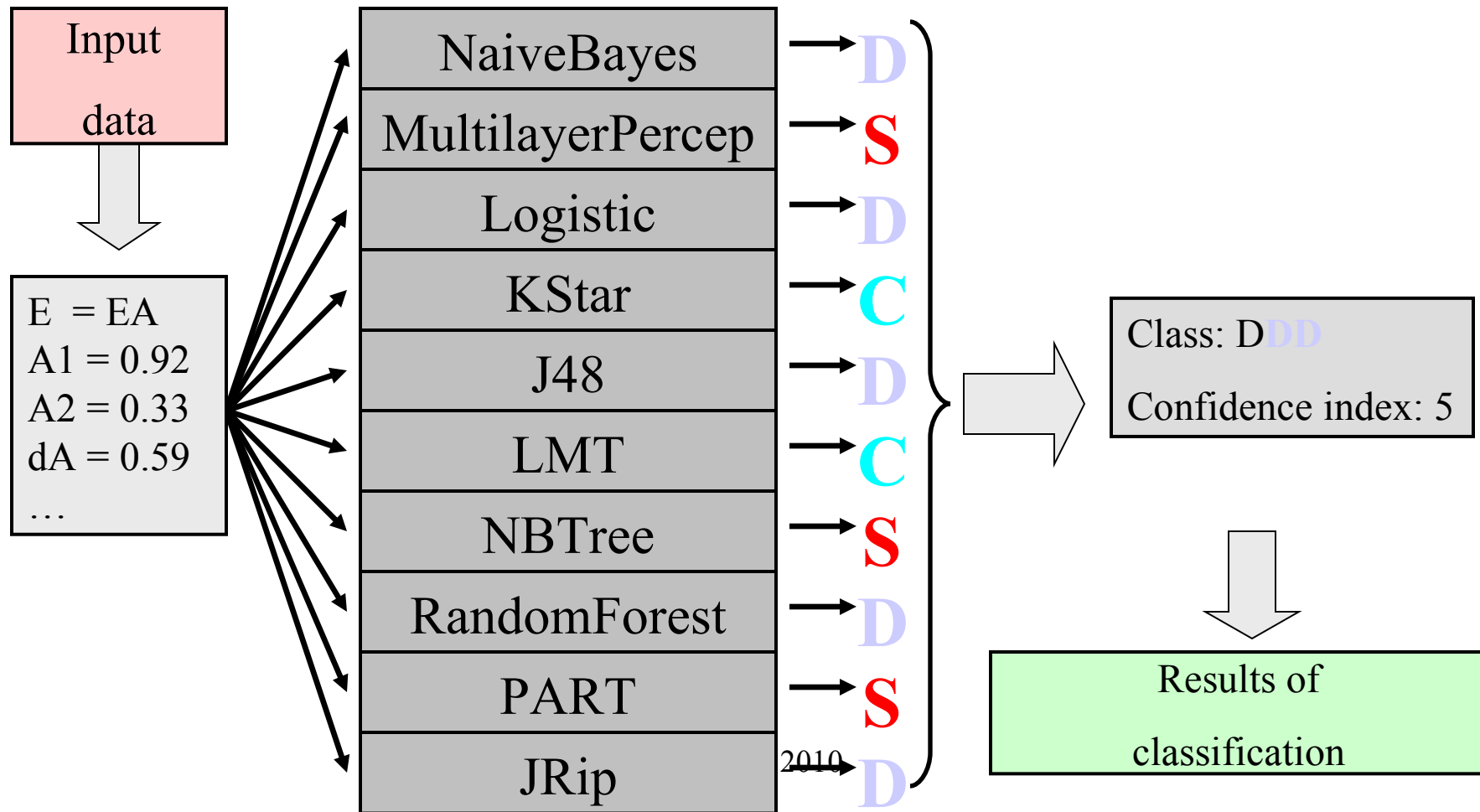
# Weka and AstroWeka

- Existing Data mining instruments include: Java Data Mining, Oracle Data Mining, MATLAB, Weka.
- Weka has been chosen for this work as an open source implementing a large number of various Data Mining algorithms. It appeared also that Weka has been planned in the UK for incorporation into AstroGrid as AstroWeka
- AstroWeka is an extension of graphical interface of Weka providing for processing of data in the AstroGrid MySpace and for data exchange with other applications by means of the interface of PLASTIC (PLatform for AStronomical Tool InterConnection).
- AstroWeka is implemented as a separate application, not as an AstroGrid service.

# Algorithms selected for EBs

- **Bayes Classifiers**
  - NaiveBayes
- **Functional Classifiers**
  - MultilayerPerceptron
  - Logistic
- **Pattern matching Classifiers**
  - KStar
- **Decision trees**
  - J48
  - LMT
  - NBTree
  - RandomForest
- **Decision rules**
  - PART
  - JRip

# Schema of Ensemble Use for Classification



# Result of classification

- Applying Ensemble Weka in AstroGrid 5514 binaries were classified providing the following star classes distribution:
  - C - 852
    - CB - 89
    - CBF - 74
    - CBV - 149
    - CE - 15
    - CG - 1
    - CW - 84
    - CWA - 427
    - CWW - 331
  - S - 547
    - S2C - 3
    - SA - 1902
    - SC - 1
    - SH - 13
  - D - 553
    - DG - 41
    - DM - 422
    - DR - 10

As a threshold for the confidence index 7 has been chosen

# Weka approach: conclusions

- Ensembled Weka installed at the AstroGrid installation of IPI RAS and published at the AstroGrid registries. Therefore, Ensembled Weka is ready for use.
- Ensembled Weka has been checked by solving of eclipsing binaries classification problem.
- It is planned to extend Ensembled Weka for solving clustering and regression kinds of Data Mining problems applying ensembles of algorithms.

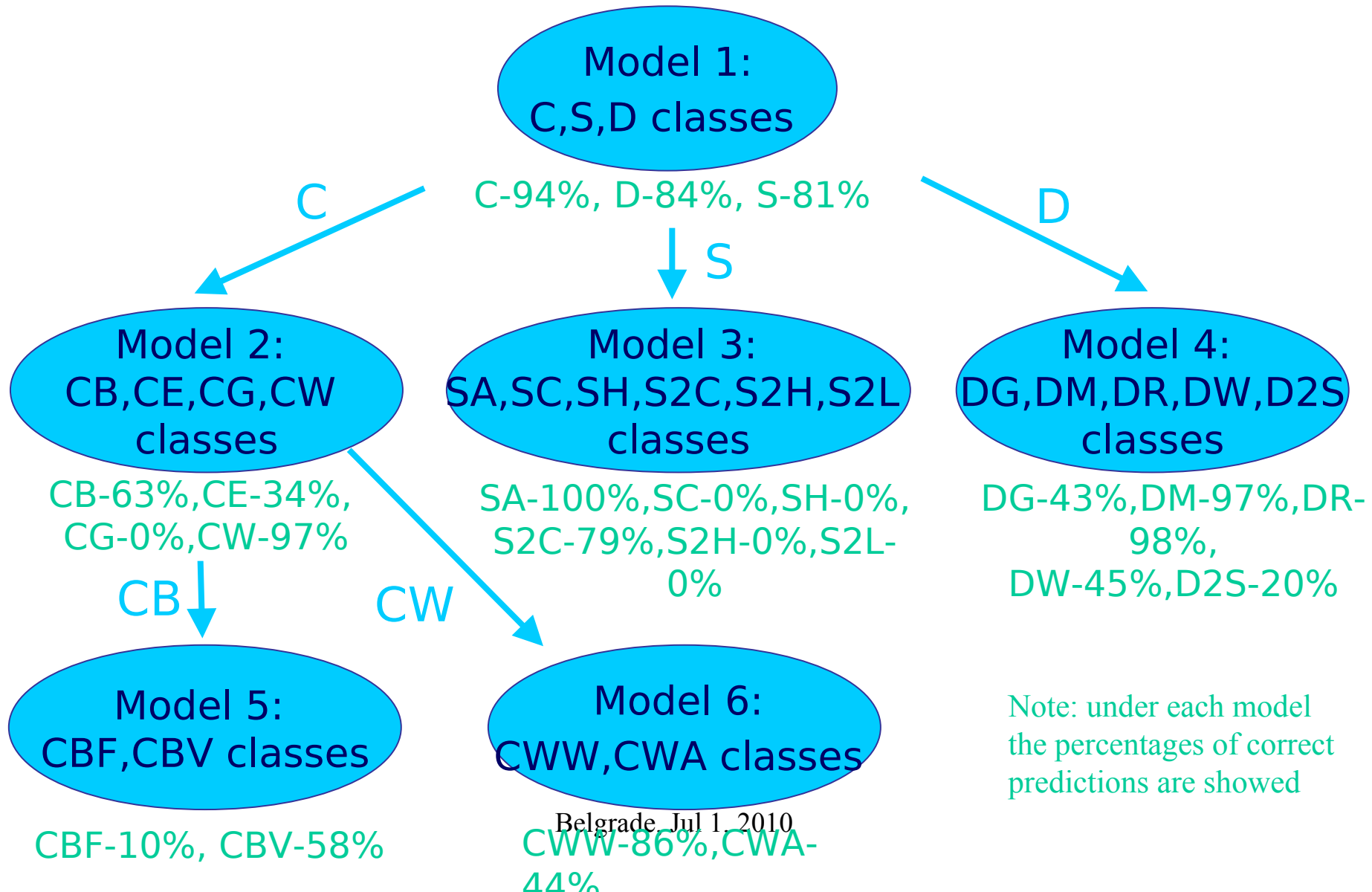
# Oracle approach. 1

- Classification of eclipsing binaries was performed by the Support Vector Machine (SVM) algorithm. The SVM is a powerful machine learning method for the statistical classification of data used in various application domains. The main advantages of this method are high classification accuracy, robustness against overfitting and well-founded theory. Another reasons for the application of SVM for classification of eclipsing binaries are the large number of the eclipsing binaries parameters most of which are numerical.

# Oracle approach. 2

- The training set that was used for building of the classification model consisted of 1162 binaries having the manually assigned classes. Due to existing hierarchy in the considered classification scheme of the eclipsing binaries we constructed six classification models by one on each node of the hierarchy classes tree. All classification models were built using the Oracle's realization of the SVM algorithm that is part of the Oracle 10g Data Mining module.

# Hierarchy of the classification models, testing results



# Oracle approach: conclusions

- The testing of the resulted classification models was accomplished by 10-fold cross validation approach. It was highest for the first hierarchy level of the classification scheme (contact system, semi-detached system and detached system) - about 90% - and less precise for other levels.

# Astrophysical applications

- Fine structure of the MLR for intermediate masses. Evolution, metallicity, rotation effects.
- Subgiant (IV) sequence on HRD and interactive binaries population.
- Status of some types of EB (essentially contact and near-contact) in evolutionary scheme.
- Study of marginal systems.

# Acknowledgements

- Dana Kovaleva, Sergej Sichevskij, Leonid Kalinichenko, Marat Kazanov for collaboration and valuable comments
- Russian Foundation for Fundamental Researches for financial support (09-02-00520, 10-02-00426)
- The school in Astrodinformatcs – Virtual Observatory organizers for invitation
- Audience for your attention

**Bibliography:**

- Malkov et al. 2010, MNRAS 401, 695
- Malkov 2007, MNRAS 382, 1073
- Malkov et al. 2007, AA 465, 549
- Malkov et al. 2006, AA 446, 785