Electronic Supplementary Information for article:

Pattern recognition as a new strategy in high-resolution spectroscopy: Application to methanol OH-stretch overtones.

Jozef Rakovský,[a] Vít Svoboda,[a,b] Veronika Horká-Zelenková[a,c], and Ondrej Votava,*[a]

[a] *J. Heyrovský Institute of Physical Chemistry of the ASCR, v.v.i., Dolejškova 2155/3, 182 23 Prague 8, Czech Republic*
[b] *Laboratory of Physical Chemistry, ETH Zürich, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland*
[c] *ASCR, Institute of Physics, v.v.i., Na Slovance 1999/2, 182 21 Prague 8, Czech Republic*
* *E-mail: ondrej.votava@jh-inst.cas.cz*

1 Measured spectrum

Figure 1 shows a zoomed view of the methanol spectrum between 7200.6 cm^{-1} and 7200.9 cm^{-1} measured at 19 K. Despite this low temperature, the spectrum is highly congested with ro-vibrational lines (26 lines in 0.3 cm^{-1}). The figure demonstrates the complexity of overtone spectra which makes any traditional assignment procedures impossible.

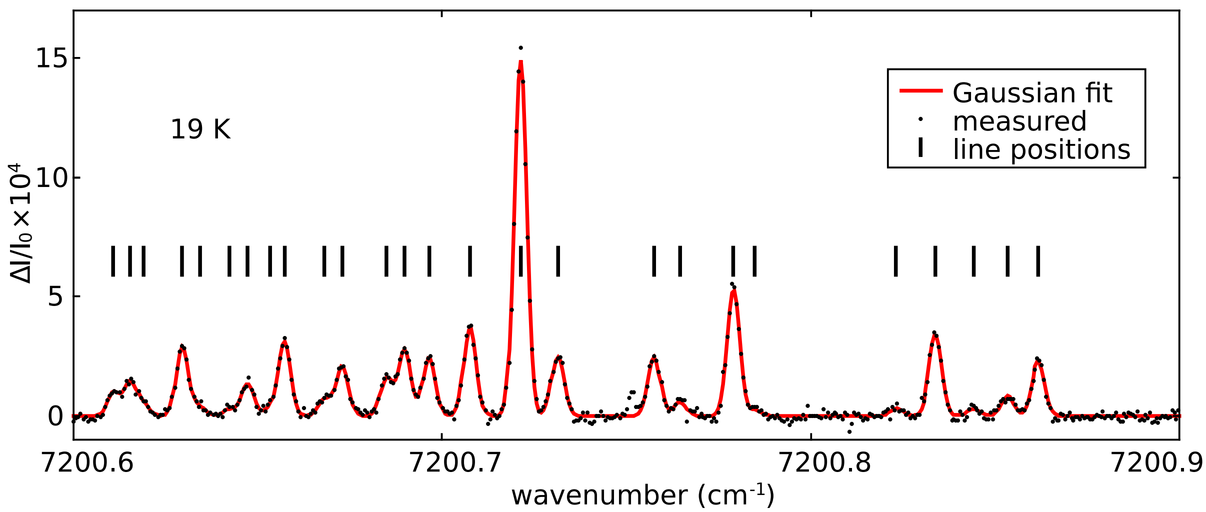

Figure 1 A detail of the methanol spectrum measured between 7200.6 cm^{-1} and 7200.9 cm^{-1} at 19 K. Measured data (black points) and multi-Gaussian fit (red solid line) are shown together with line positions marked as black vertical lines. In total, 26 lines are fitted in this portion of the spectrum. Line widths are fixed to a value corresponding to the temperature of 19 K.

2 Graphical representation of methanol multiplets

Figure 2 shows a schematic diagram of rotational states for both A and E symmetry components of methanol. Each rotational state is represented by a black horizontal line. All states are organized according to their respective J and K rotational quantum numbers into a matrix-like form with J decreasing along the rows and K increasing along the columns. For $K > 0$, the A states are additionally split into doublets denoted as $A+$ and $A-$, respectively, and displayed in the diagram as two closely spaced lines. Since the rotational structure is the same in both upper and lower vibrational states, they can be represented by the same diagram.

The well-known rotational selection rules for methanol enable to visualize all allowed transition from a particular lower ro-vibrational state (with given J'' and K'') using a 'selection rule stencil' represented by red dots in Figure 2. Methanol multiplets have between 2 and 9 transitions for A component, and between 3 to 8 transitions for E component, respectively. When a given stencil is overlaid with the diagram for the upper ro-vibrational state and centered on a corresponding upper rotational state with same quantum numbers J' and K' (thick horizontal line), all allowed transitions can be easily deduced. If a red dot falls on a particular rotational state in the diagram, there is an allowed transition between the respective lower and upper rotational state. The transition may or may not be observed in a measured spectrum depending on several factors such as cross-section and experimental signal-to-noise ratio.

A lot of information about the multiplets can be deduced from the graphical representation. For example, some transitions are unique to a particular multiplet. This allows to unambiguously assign a multiplet even if not all lines from the multiplet are observed in the spectrum. These unique lines tend to lie in the corners of the stencil. Furthermore, shared transitions between two different multiplets can be visualized as well using two stencils. Figure 3 shows four real scenarios. Two stencils are highlighted in red circles and black rectangles, respectively. In these scenarios, if a particular line is found in the measured spectrum, the circle or rectangle is filled otherwise the mark is empty. For example, in scenario A, two stencils represent upper state multiplets $(3,1,A+)$ and $(3,2,A+)$ sharing mutually six lower rotational states and six ro-vibrational transitions. Similar results can be obtained for the remaining scenarios B - D. It can be shown, that two different multiplets can share a maximum of 6 and 4 lower rotational states for A and E component, respectively.

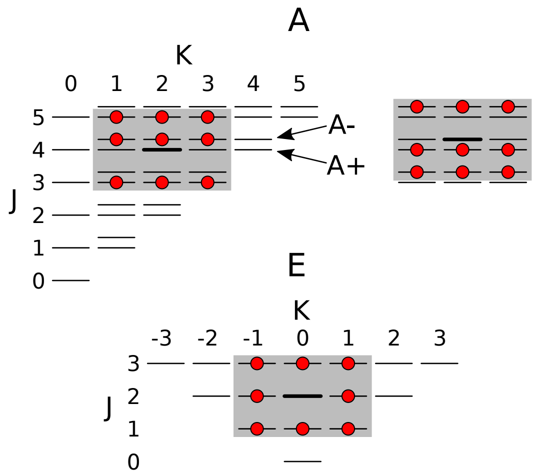

Figure 2 A graphical representation of multiplets with the highest possible number of lines, showing three 'selection rule stencils' (for $A+$, $A-$ and E symmetry states, respectively) suitable for the generation of any methanol multiplet. Thin lines represent energy states, thick lines mark the selected upper state and circles indicated all lower states of the multiplet for a given upper state.

3 Reduced energy plots

Every unique multiplet determines the energy and rotational assignment of a single excited ro-vibrational state. Vibrational assignments however can only be inferred if a regular progressions of rotational states are observed. To visualize such progressions, it is useful to collect the identified upper rotational states in the form of reduced energy plots, as discussed in the section 5.2 of the paper. Reduced energies of rotational states belonging to the same vibration should vary only slightly as a function of J' and thus the progressions are observed as almost horizontal lines across the reduced energy plots. Here we discuss, how the reduced energy plots have been used to determine correct rotational assignments for non-unique patterns observed in the measured spectrum. First, we construct the plots using only the patterns that uniquely determine their multiplets. This creates a back-bone of reliably assigned states, where the major progressions can be easily followed. Then, the correct rotational assignments for non-unique patterns, where several rotational assignments are possible, are determined.

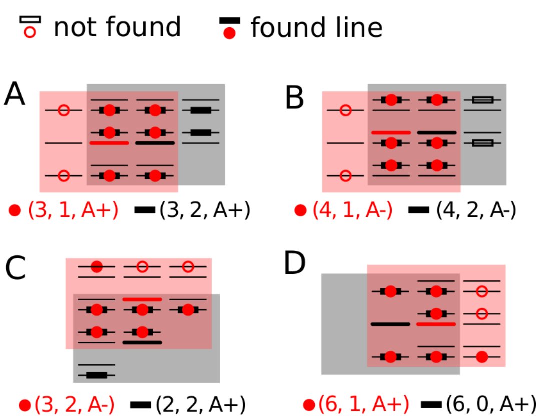

Figure 3 A graphical representation of four non-unique pattern scenarios (A - D) considered in the analysis of the methanol spectrum. The scenarios demonstrate possible difficulties during the multiplet assignments due to mutual sharing of lower rotational states between different multiplets (here represented in red and black). The shared states are located in the intersection between two rectangular boxes representing individual multiplets. Full labels represent lines identified in the spectrum, while unobserved transitions are shown as empty symbols.

For illustration, several such cases are shown in Figure 3. For comparison panel A shows situation where two multiplets share 6 transitions, however two distinguishing transitions are observed allowing unique assignment. In contrast panel B is a case where none of the distinguishing lines have been observed and thus the pattern recognition method cannot distinguish between the two possibilities. In this case, the reduced energy plot allows to make the final assignment by following the regular progressions. As pointed out in the Figure 4, where based on this additional information, the $(4, 2, A-)$ multiplet is assigned as the correct one.

Panels C and D show situation, where an extra line is assigned to a pattern due to a accidental coincidence making the assignment either ambiguous (C) or incorrect (D). Specifically in panel C, $(3, 2, A-)$ and $(2, 2, A+)$ multiplets are show sharing 5 lower rotational states. In addition, two other lines are found each indicating different assignments. Thus the pattern recognition method cannot distinguish these two multiplets from each other. However, $(3, 2, A-)$ multiplet cannot be the correct assignment since it does not follow any of the previously identified progression, see Figure 4. Therefore, we can conclude that the correct assignment is $(2,2,A+)$ multiplet.

Finally in panel D, $(6,1,A+)$ and $(6,0,A+)$ multiplets are shown sharing 5 lower rotational states. Moreover, $(6,1,A+)$ multiplet has one additional line found in the measured spectrum. Therefore, the patter recognition method would assign this multiplet as an unique one. However, this assignment is incorrect due to an accidental coincidence in the pattern. This has been again inferred from the reduced energy plot in Figure 4, where this assignment as $(6,1,A+)$ multiplet does not fall into any previously identified progression. Thus, the correct assignment is $(6,0,A+)$ multiplet as confirmed by the progression in the reduced energy plot.

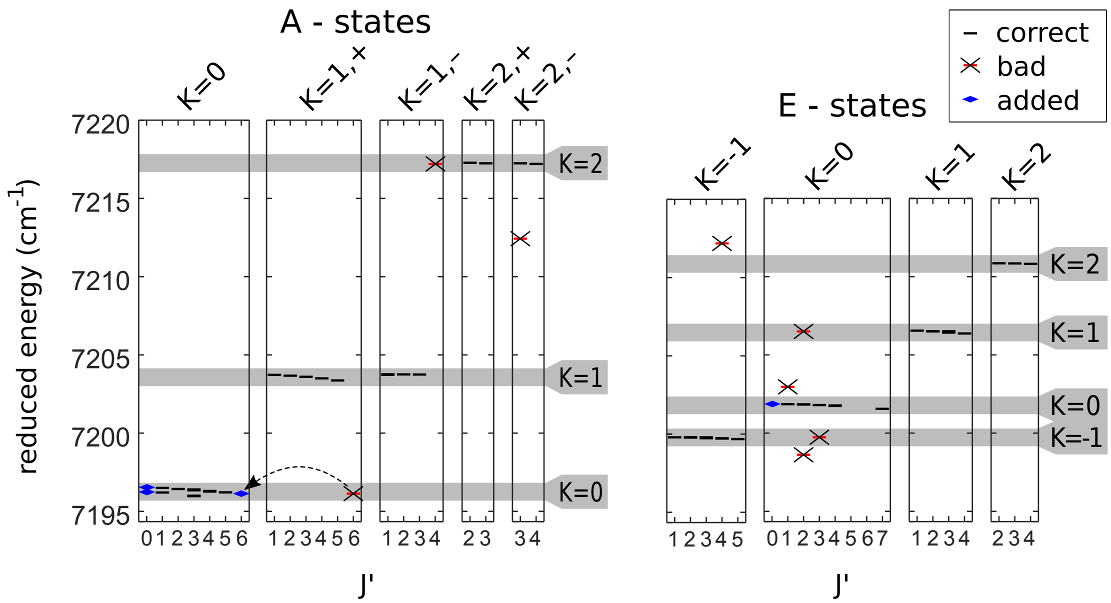

Figure 4 The reduced energy plots for final set of all identified multiplets. The black and red dashes mark the multiplets fulfilling all defined criteria resulting from automated data manipulation. The blue diamonds mark the multiplets added later manually. Each plot is associated with different K quantum number and $A\pm$ or E symmetry. Horizontal, 1 cm^{-1} wide gray bars highlight progressions labeled by their relation to quantum number K.

4 Complete list of reliably identified multiplets

Table 1 summarizes all assigned multiplets in the methanol spectrum in the first OH-stretch overtone region, $2\nu_{\mathrm{OH}}$, between 7170 cm^{-1} and 7220 cm^{-1}. In total, all multiplets represent 37 upper rotational states. The table is organized as follows. Each multiplet is associated with its upper rotational state (first column). Some of these multiplets appear more than once in the spectrum (1 - 4 times). Their respective upper rotational energies are listed in the second column. The remaining columns represent line positions in the measured spectrum (bottom row), used in the identification of the respective multiplets, together with their lower rotational state energies and quantum numbers (top row).

For example, upper rotational state $(0,0,A+)$ is listed in the first entry of Table 1. Altogether, two upper energies (7196.5368 cm^{-1} and 7196.2342 cm^{-1}) have been identified suggesting at least two upper vibrational states in the measured spectral range. Each state (multiplet) has two transitions in the spectrum, originating from $(1,0,A+)$ and $(1,1,A+)$ lower rotational states with energies of 1.6135 cm^{-1} and 11.7049 cm^{-1}, respectively.

Table 1 A final list of all assigned multiplets. Information on how to read the table is given in Section 4.

upper state	E' (cm^{-1})	lower state, E" (cm^{-1}), lines(cm^{-1})								
(0, 0, A+)		(1, 0, A+) 1.6135	(1, 1, A+) 11.7049							
	7196.5368	7194.9234	7184.8318							
	7196.2342	7194.621	7184.529							
(1, 0, A+)		(0, 0, A+) 0	(1, 1, A−) 11.7328	(2, 0, A+) 4.8405	(2, 1, A+) 14.9043					
	7198.1141	7198.1143	7186.3813	7193.2738	7183.2096					
	7197.821	7197.8212	7186.0882	7192.9806	7182.9163					
(1, 1, A−)		(1, 0, A+) 1.6135	(1, 1, A+) 11.7049	(2, 1, A−) 14.9878	(2, 2, A−) 31.049					
	7205.3774	7203.7636	7193.6725	7190.3897	7174.3287					
	7205.3451	7203.7319	7193.6402	7190.3571	7174.2962					
(1, 1, A+)		(0, 0, A+) 0	(1, 1, A−) 11.7328	(2, 0, A+) 4.8405	(2, 1, A+) 14.9043	(2, 2, A+) 31.049				
	7205.3378	7205.3375	7193.605	7200.4974	7190.4335	7174.2889				
(2, 0, A+)		(1, 0, A+) 1.6135	(1, 1, A+) 11.7049	(2, 1, A−) 14.9878	(3, 0, A+) 9.6806	(3, 1, A+) 19.7031				
	7201.2804	7199.6671	7189.5756	7186.2927	7191.5998	7181.577				
(2, 1, A−)		(1, 1, A−) 11.7328	(2, 0, A+) 4.8405	(2, 1, A+) 14.9043	(2, 2, A+) 31.049	(3, 1, A−) 19.8701	(3, 2, A−) 35.8898			
	7208.6037	7196.8708	7203.7636	7193.6991	7177.555	7188.7332	7172.7138			
(2, 1, A+)		(1, 0, A+) 1.6135	(1, 1, A+) 11.7049	(2, 1, A−) 14.9878	(2, 2, A−) 31.049	(3, 0, A+) 9.6806	(3, 1, A+) 19.7031	(3, 2, A+) 35.8902		
	7208.521	7206.9076	7196.816	7193.5332	7177.4723	7198.8405	7188.8177	7172.6308		
(2, 2, A+)		(1, 1, A+) 11.7049	(2, 1, A−) 14.9878	(2, 2, A−) 31.049	(3, 1, A+) 19.7031	(3, 2, A+) 35.8902	(3, 3, A+) 44.2928			
	7222.1281	7210.4231	7207.1404	7191.0788	7202.4257	7186.2377	7177.8352			
(3, 0, A+)		(2, 0, A+) 4.8405	(2, 1, A+) 14.9043	(3, 1, A−) 19.8701	(4, 0, A+) 16.1335	(4, 1, A+) 26.1012				
	7206.1036	7201.2634	7191.1994	7186.2333	7189.9699	7180.0025				
	7206.0452	7201.2049	7191.1407	7186.175	7189.9115	7179.9441				
	7205.7019	7200.8617	7190.7976	7185.8318	7189.5684	7179.6005				
	7205.6741	7200.8338	7190.7698	7185.8038	7189.5403	7179.5733				
(3, 1, A−)		(2, 1, A−) 14.9878	(2, 2, A−) 31.049	(3, 0, A+) 9.6806	(3, 1, A+) 19.7031	(3, 2, A+) 35.8902	(4, 1, A−) 26.3795	(4, 2, A−) 42.3438		
	7213.4458	7198.4581	–	7203.7656	7193.7429	7177.555	7187.0661	7171.1019		
(3, 1, A+)		(2, 0, A+) 4.8405	(2, 1, A+) 14.9043	(2, 2, A+) 31.049	(3, 1, A−) 19.8701	(3, 2, A−) 35.8898	(4, 0, A+) 16.1335	(4, 1, A+) 26.1012	(4, 2, A+) 42.345	
	7213.3008	7208.4604	7198.3965	–	7193.4306	7177.4114	7197.1674	7187.1992	7170.956	
(3, 2, A−)		(2, 1, A−) 14.9878	(2, 2, A−) 31.049	(3, 1, A+) 19.7031	(3, 2, A+) 35.8902	(3, 3, A+) 44.2928	(4, 1, A−) 26.3795	(4, 2, A−) 42.3438	(4, 3, A−) 50.7463	
	7226.9458	7211.9582	7195.8968	7207.2427	7191.0558	7182.653	7200.5666	7184.6011	7176.1996	
(3, 2, A+)		(2, 1, A+) 14.9043	(2, 2, A+) 31.049	(3, 1, A−) 19.8701	(3, 2, A−) 35.8898	(3, 3, A−) 44.2928	(4, 1, A+) 26.1012	(4, 2, A+) 42.345	(4, 3, A+) 50.7463	
	7226.9457	7212.0414	7195.8968	7207.0755	7191.0558	7182.653	7200.8442	7184.6011	7176.1996	
(4, 0, A+)		(3, 0, A+) 9.6806	(3, 1, A+) 19.7031	(4, 1, A−) 26.3795	(5, 0, A+) 24.1988	(5, 1, A+) 34.0983				
	7212.4727	7202.7924	7192.7695	7186.093	7188.2736	7178.3746				
	7212.4342	7202.7539	7192.731	7186.0544	7188.2353	7178.3362				
(4, 1, A+)		(3, 0, A+) 9.6806	(3, 1, A+) 19.7031	(3, 2, A+) 35.8902	(4, 1, A−) 26.3795	(4, 2, A−) 42.3438	(5, 0, A+) 24.1988	(5, 1, A+) 34.0983	(5, 2, A+) 50.4135	
	7219.6698	7209.9893	7199.9668	7183.7789	7193.2903	7177.3262	7195.471	7185.5714	–	
(4, 2, A−)		(3, 1, A−) 19.8701	(3, 2, A−) 35.8898	(3, 3, A−) 44.2928	(4, 1, A+) 26.1012	(4, 2, A+) 42.345	(4, 3, A+) 50.7463	(5, 1, A−) 34.5156	(5, 2, A−) 50.4108	(5, 3, A−) 58.813
	7233.3726	7213.5026	7197.4829	–	7207.2715	7191.0272	–	–	7182.9611	7174.5603
(5, 0, A+)		(4, 0, A+) 16.1335	(4, 1, A+) 26.1012	(5, 1, A−) 34.5156	(6, 0, A+) 33.8759	(6, 1, A+) 43.694				
	7220.4703	7204.3368	7194.3692	7185.9544	7186.5942	7176.7764				
(5, 1, A+)		(4, 0, A+) 16.1335	(4, 1, A+) 26.1012	(4, 2, A+) 42.345	(5, 1, A−) 34.5156	(5, 2, A−) 50.4108	(6, 0, A+) 33.8759	(6, 1, A+) 43.694	(6, 2, A+) 60.0957	
	7227.6219	7211.4884	7201.5208	7185.2764	7193.1061	7177.2118	7193.7462	7183.9275	–	
(6, 0, A+)		(5, 0, A+) 24.1988	(5, 1, A+) 34.0983	(6, 1, A−) 44.2782	(7, 0, A+) 45.164	(7, 1, A+) 54.8878				
	7230.0902	7205.8915	7195.992	7185.8117	7184.926	7175.2027				
(0, 0, E)		(1, -1, E) 5.4897	(1, 0, E) 10.7357	(1, 1, E) 16.2412						
	7201.8633	7196.3737	7191.1274	7185.622						

upper state	E' (cm^{-1})	lower state, E" (cm^{-1}), lines(cm^{-1})							
(1, -1, E)		(0, 0, E) 9.122	(1, 0, E) 10.7357	(2, -2, E) 22.841	(2, -1, E) 8.7166	(2, 0, E) 13.9628			
	7201.4086	7192.2866	7190.6729	7178.5679	7192.6919	7187.4455			
(1, 0, E)		(0, 0, E) 9.122	(1, -1, E) 5.4897	(1, 1, E) 16.2412	(2, -1, E) 8.7166	(2, 0, E) 13.9628	(2, 1, E) 19.4686		
	7203.4675	7194.3456	7197.9778	7187.2262	7194.7509	7189.5047	7183.9986		
(1, 1, E)		(0, 0, E) 9.122	(1, 0, E) 10.7357	(2, 0, E) 13.9628	(2, 1, E) 19.4686	(2, 2, E) 20.3003			
	7208.1644	7199.0426	7197.4289	7194.2015	7188.6955	7187.8639			
(2, -1, E)		(1, -1, E) 5.4897	(1, 0, E) 10.7357	(2, -2, E) 22.841	(2, 0, E) 13.9628	(3, -2, E) 27.6819	(3, -1, E) 13.5565	(3, 0, E) 18.8026	
	7204.6236	7199.134	7193.8879	7181.7824	7190.6609	7176.9419	7191.0668	7185.8208	
	7204.6503	7199.1607	7193.9146	–	7190.6878	7176.9687	7191.0937	7185.8474	
(2, 0, E)		(1, -1, E) 5.4897	(1, 0, E) 10.7357	(1, 1, E) 16.2412	(2, -1, E) 8.7166	(2, 1, E) 19.4686	(3, -1, E) 13.5565	(3, 0, E) 18.8026	(3, 1, E) 24.3097
	7206.6911	7201.2009	7195.9557	7190.4504	7197.9744	7187.2224	7193.1348	7187.8886	7182.3812
	7206.6765	7201.1873	7195.941	–	7197.9601	7187.2071	7193.1203	7187.8739	7182.3665
(2, 1, E)		(1, 0, E) 10.7357	(1, 1, E) 16.2412	(2, 0, E) 13.9628	(2, 2, E) 20.3003	(3, 0, E) 18.8026	(3, 1, E) 24.3097	(3, 2, E) 25.1412	
	7211.3512	7200.6153	7195.1097	7197.3882	7191.051	7192.5487	7187.041	7186.2107	
	7211.3652	7200.6295	7195.1241	7197.4025	–	7192.5628	7187.0553	7186.2242	
(2, 2, E)		(1, 1, E) 16.2412	(2, 1, E) 19.4686	(3, 1, E) 24.3097	(3, 2, E) 25.1412	(3, 3, E) 42.8417			
	7215.7057	7199.4644	7196.2369	7191.3964	7190.5645	7172.864			
(3, -1, E)		(2, -2, E) 22.841	(2, -1, E) 8.7166	(2, 0, E) 13.9628	(3, -2, E) 27.6819	(3, 0, E) 18.8026	(4, -2, E) 34.1368	(4, -1, E) 20.0091	(4, 0, E) 25.2542
	7209.4955	–	7200.7791	7195.5326	7181.8132	7190.693	7175.3591	7189.4865	–
	7209.4377	–	7200.7212	7195.475	–	7190.6349	7175.3013	7189.4286	7184.1833
(3, 0, E)		(2, -1, E) 8.7166	(2, 0, E) 13.9628	(2, 1, E) 19.4686	(3, -1, E) 13.5565	(3, 1, E) 24.3097	(4, -1, E) 20.0091	(4, 0, E) 25.2542	(4, 1, E) 30.7644
	7211.4969	7202.7807	7197.5342	7192.0284	7197.9404	7187.187	7191.488	7186.2426	7180.7324
(3, 1, E)		(2, 0, E) 13.9628	(2, 1, E) 19.4686	(2, 2, E) 20.3003	(3, 0, E) 18.8026	(3, 2, E) 25.1412	(4, 0, E) 25.2542	(4, 1, E) 30.7644	(4, 2, E) 31.5961
	7216.1235	7202.1613	7196.655	7195.8223	7197.3209	7190.9825	7190.8694	7185.359	–
	7216.2147	7202.2514	7196.7461	7195.9152	7197.4118	–	–	7185.4502	7184.6187
	7216.186	7202.2236	7196.7172	–	7197.3834	7191.0444	–	7185.4214	7184.5906
(3, 2, E)		(2, 1, E) 19.4686	(2, 2, E) 20.3003	(3, 1, E) 24.3097	(3, 3, E) 42.8417	(4, 1, E) 30.7644	(4, 2, E) 31.5961	(4, 3, E) 49.2953	
	7220.5465	7201.078	7200.2461	7196.2369	7177.7053	7189.7819	7188.9501	7171.2513	
(4, -1, E)		(3, -2, E) 27.6819	(3, -1, E) 13.5565	(3, 0, E) 18.8026	(4, -2, E) 34.1368	(4, 0, E) 25.2542	(5, -2, E) 42.2058	(5, -1, E) 28.0735	(5, 0, E) 33.3165
	7215.9029	–	7202.347	7197.1007	7181.7655	7190.6487	7173.6971	7187.8291	7182.5862
	7215.8833	–	7202.327	7197.0809	7181.7463	7190.6296	7173.6777	7187.8092	7182.5665
(4, 0, E)		(3, -1, E) 13.5565	(3, 0, E) 18.8026	(3, 1, E) 24.3097	(4, -1, E) 20.0091	(4, 1, E) 30.7644	(5, -1, E) 28.0735	(5, 0, E) 33.3165	(5, 1, E) 38.8326
	7217.9073	7204.3508	7199.1049	7193.5977	7197.8982	7187.1428	7189.8336	7184.5906	7179.0751
	7217.9397	7204.3834	7199.1375	7193.6301	7197.9309	7187.1752	7189.8653	7184.6233	–
(4, 1, E)		(3, 0, E) 18.8026	(3, 1, E) 24.3097	(3, 2, E) 25.1412	(4, 0, E) 25.2542	(4, 2, E) 31.5961	(5, 0, E) 33.3165	(5, 1, E) 38.8326	(5, 2, E) 39.6651
	7222.5383	7203.7362	7198.2287	7197.3976	7197.284	7190.9424	7189.2211	7183.7053	–
(4, 2, E)		(3, 1, E) 24.3097	(3, 2, E) 25.1412	(3, 3, E) 42.8417	(4, 1, E) 30.7644	(4, 3, E) 49.2953	(5, 1, E) 38.8326	(5, 2, E) 39.6651	(5, 3, E) 57.3624
	7226.9854	7202.6758	7201.8446	–	7196.2211	–	7188.1522	7187.3201	7169.6233
(5, -1, E)		(4, -2, E) 34.1368	(4, -1, E) 20.0091	(4, 0, E) 25.2542	(5, -2, E) 42.2058	(5, 0, E) 33.3165	(6, -2, E) 51.8894	(6, -1, E) 37.7492	(6, 0, E) 42.9878
	7223.9347	7189.7983	7203.9251	7198.6811	7181.7286	7190.6182	–	7186.1853	–
(7, 0, E)		(6, -1, E) 37.7492	(6, 0, E) 42.9878	(6, 1, E) 48.5142	(7, -1, E) 49.0352	(7, 1, E) 59.8092	(8, -1, E) 61.9304	(8, 0, E) 67.1503	(8, 1, E) 72.7174
	7246.8267	7209.0777	7203.8387	7198.3129	7197.7911	7187.0169	7184.8964	7179.6767	–