

The Need for Speed

The Need for Speed

The PDF-4+ database is designed to handle very large amounts of data and provide the user with an ability to perform extensive data mining.

The database also incorporates many sophisticated algorithms to calculate, analyze and display diffraction data.

By design, speed is sacrificed whenever there is a trade-off between speed and capability. However speed improvements are made annually and this tutorial gives many helpful hints on how to improve speed for any database product.

Speed Gobblers (Speed bumps)

- Sophisticated on-the fly calculations
- Inadequate computer hardware
- Sorts and searches producing or using very large tables

Speed Accelerators (Turbo)

- Smaller data sets
- Targeted and selective (smaller) searches producing small tables

Speed Basics

Preferred Computer Specifications

	PDF-4+ 2009	WebPDF-4+ 2009†	Minerals 2009	Organics 2009	Organics 2010
Intel® Core™ 2 Duo processors or equivalent*	✓	✓	✓	✓	✓
Supported operating systems:					
Microsoft® Windows® 2000 SP4	✓	✓	✓	✓	✓
Microsoft® Windows® XP® Professional SP3	✓	✓	✓	✓(SP2)	✓
Microsoft® Windows® Vista™ (32-bit) Business Edition SP1	✓	✓	✓	✓	✓
NTFS file system required	✓	✓	✓	✓	✓
System Memory**	2 GB	2 GB	2 GB	2 GB	2 GB
Installation requires free hard-disk space	6 GB	0.4 GB	1 GB	14 GB	16 GB

† Requires USB port, web access, internet browser IE 6.0 or newer (or Firefox 2.0 or newer), and ability to open ports through any existing firewalls. Browser must allow Java™ and scripts to execute. Connectivity speed should be at least 1.54 Mbps with 3.0 Mbps or higher preferred.

* Minimum specification for processor is Pentium IV or equivalent.

**Minimum system memory specification is 1 GB for Windows® 2000 or XP®, 2 GB for Vista®.

PDF-2 Release 2009	DDView/Sieve with PDF-2
✓	✓
✓	✓
✓	✓
✓	✓
2 GB	2 GB
2.5 GB	3 GB

The processor and system memory control the overall speed of the database and searches that use the database. If you exceed the specifications the system will work faster, if you do not meet the specifications certain operations will work very slowly or not work at all.

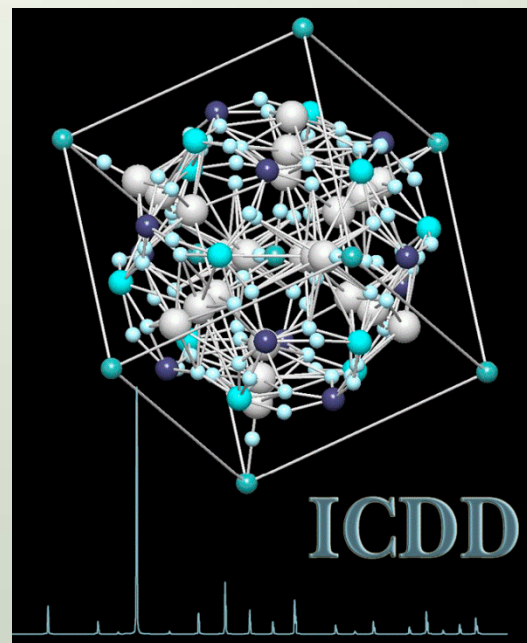
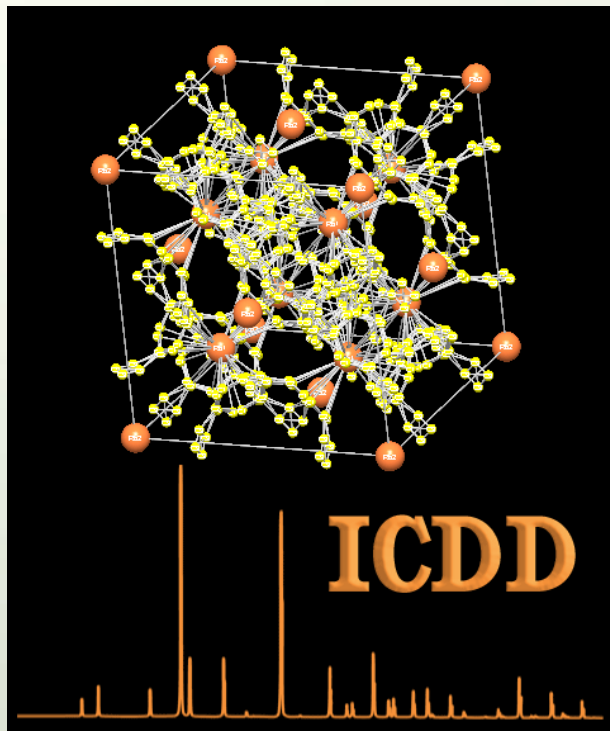
Search Speeds

There are several search speeds given in this presentation. These were given to present the reader with a concept of relative speed for several types of searches.

Your search times may vary depending on the exact computer specifications of your PC.

The PC used in these tests was 32-bit with a Vista™ operating system 2 GB RAM and 2.00 GHz dual core processor.

How fast is this spinning ?



The diffraction patterns should be smooth and continuously drawn.



The shark should swim in a full circle in under one second.

Speed Basics

- All Release 2006 and later PDF-2 and PDF-4+ databases use Sybase as a database platform and use JAVA as a software platform.
- iAnywhere, the producer of Sybase, ICDD, and JAVA all continuously upgrade their systems to improve search speeds. You should see faster speeds with each product release, if you do a comparative analysis.

Database Search Options

Search

Search Options

Release 2009

PDF-4+ 53 Searches, 291,440 Data Sets

PDF-2 49 Searches, 218,610 Data Sets

PDF-4/Organics 48 Searches, 370,844 Data Sets

Searches can be combined.

Display

PDF #	QM	Chemical Formula	Compound Name	D1	D2	D3	SYS
00-001-0108	O	C6 Fe K3 O12 · 3 H2 O	Potassium Iron Oxalate Hydrate	6.900000	3.610000	2.180000	X
00-001-0117	O	Te O2	Tellurium Oxide	6.800000	2.820000	3.090000	X
00-001-0119	O	Fe3 (As O4)2 · 8 H2 O	Iron Arsenate Hydrate	6.700000	3.990000	3.700000	X
00-001-0122	I	Mg3 (P O4)2 · 8 H2 O	Magnesium Phosphate Hydrate Oxide	6.700000	2.940000	2.690000	M
00-001-0127	O	Na N H4 H P O4 · 4 H2 O	Ammonium Sodium Phosphate Hydrate Oxide	6.600000	2.890000	4.600000	X
00-001-0130	I	Mg C O3 · 3 H2 O	Magnesium Carbonate Hydrate Oxide	6.500000	3.860000	2.610000	O
00-001-0136	O	Fe +3 O (O H)	Iron Oxide Hydrate	6.350000	3.300000	2.470000	X
00-001-0153	O	Fe Cl3 · 6 H2 O	Iron Chloride Hydrate	6.000000	3.140000	2.760000	X
00-001-0172	B	Mn Cl2	Manganese Chloride	5.900000	1.840000	2.580000	R
00-001-0207	O	Na2 S O4 · 10 H2 O	Sodium Sulfate Hydrate	5.500000	3.220000	3.100000	X
00-001-0217	I	Cu +2 Cl2 · 2 H2 O	Copper Chloride Hydrate	5.400000	4.030000	2.630000	O
00-001-0229	O	Ca (N O3)2 · 4 H2 O	Calcium Nitrate Hydrate	5.200000	7.800000	2.810000	X
00-001-0256	I	Fe +2 S O4 · 7 H2 O	Iron Sulfate Hydrate	4.900000	3.780000	3.230000	M
00-001-0263	I	Al2 O3 · 3 H2 O	Aluminum Oxide Hydrate	4.850000	4.370000	2.460000	M
00-001-0264	O	Al2 O3 · 3 H2 O	Aluminum Oxide Hydrate	4.850000	4.360000	2.450000	X
00-001-0265	I	Al2 O3 · 3 H2 O	Aluminum Oxide Hydrate	4.850000	1.850000	4.340000	M
00-001-0266	O	Al2 O3 · x H2 O	Aluminum Oxide Hydrate	4.850000	4.340000	2.450000	X
00-001-0273	I	As2 S3	Arsenic Sulfur	4.820000	2.700000	4.000000	M
00-001-0274	O	Cu2 +2 (O H) P O4	Copper Phosphate Hydroxide	4.810000	2.630000	2.910000	X
00-001-0287	O	Al2 O3 · 3 H2 O	Aluminum Oxide Hydrate	4.750000	4.370000	2.220000	X
00-001-0301	B	Cu S O4 · 5 H2 O	Copper Sulfur Hydrate Oxide	4.700000	4.000000	3.700000	A

Search Description
{Database (ICDD (00))} And {Subfile/Subclass (Mineral Related)}

Display Options

PDF-4+ 85 Display Fields

PDF-2 24 Display Fields

Default page displays 8 fields. This can be reduced to 1 or expanded to all fields.

Search Speed – Size of the Search

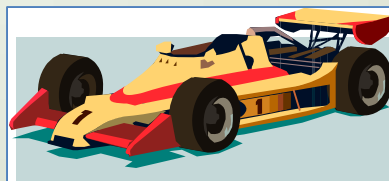
The larger the search, the slower the speed.



Display all fields (85) on all Inorganic materials

[85 X 262,365 = **22,301,025 Fields** in the table]

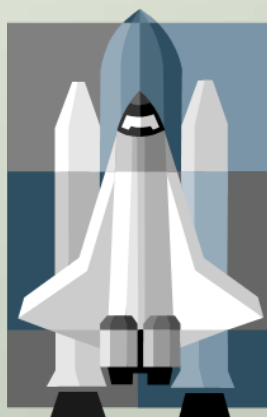
Search takes a minute



Display default fields (8) on all Inorganic Materials

[8 x 262,365 = **2,098,920 Fields** in the table]

Search takes ~ 23.8 Seconds



Display PDF#, chemical and mineral name for all Zeolites

[3 x 3,155 = **9,465 Fields** in the table]

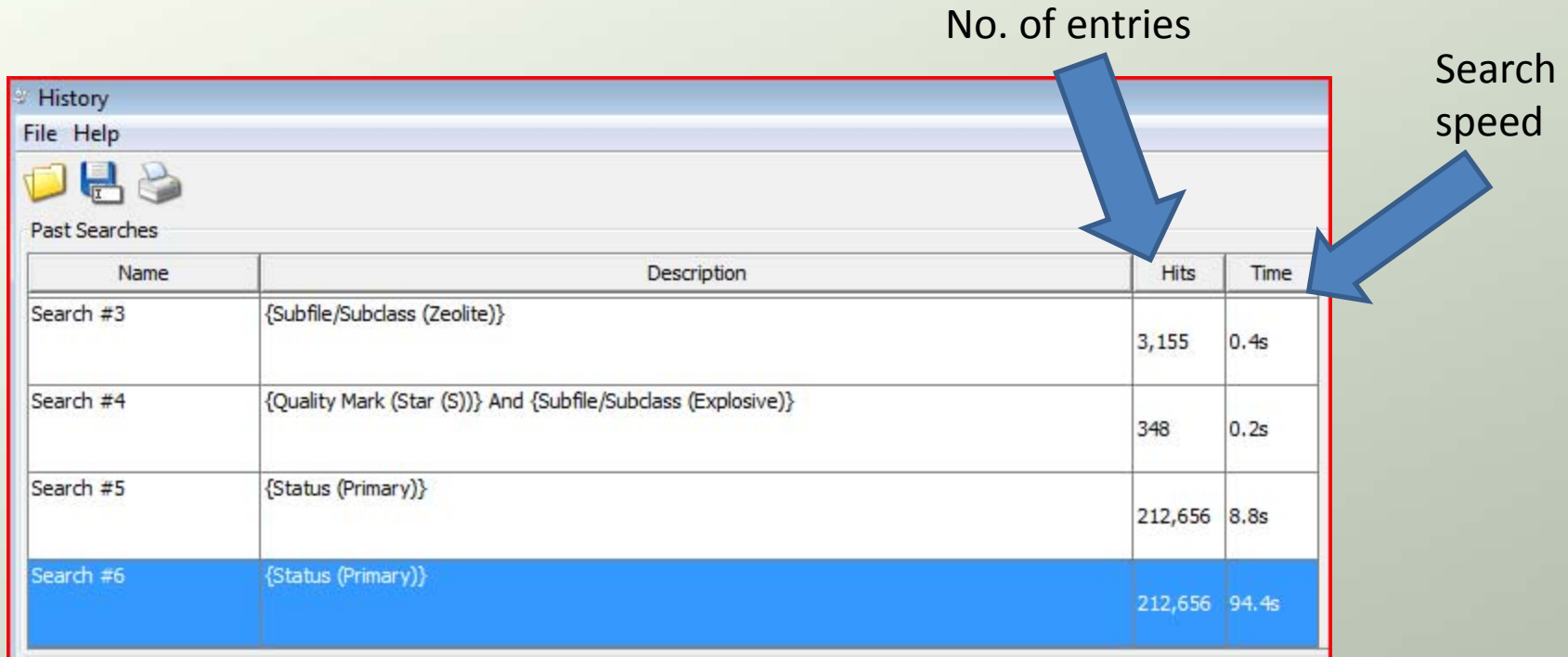
Search takes ~ Blink of the eye (0.4 Seconds)

History

The history panel tracks the search speed.

No. of entries

Search speed



Name	Description	Hits	Time
Search #3	{Subfile/Subclass (Zeolite)}	3,155	0.4s
Search #4	{Quality Mark (Star (S))} And {Subfile/Subclass (Explosive)}	348	0.2s
Search #5	{Status (Primary)}	212,656	8.8s
Search #6	{Status (Primary)}	212,656	94.4s

Searches 3 to 5, displayed 3 fields for each entry so the time is directly related to the number of entries “hit”. Search 6 at 94.4 seconds was identical to search 5 (8.8 seconds), except that 44 fields were displayed instead of 3.



Large Calculations Slow Things Down

Integral Index

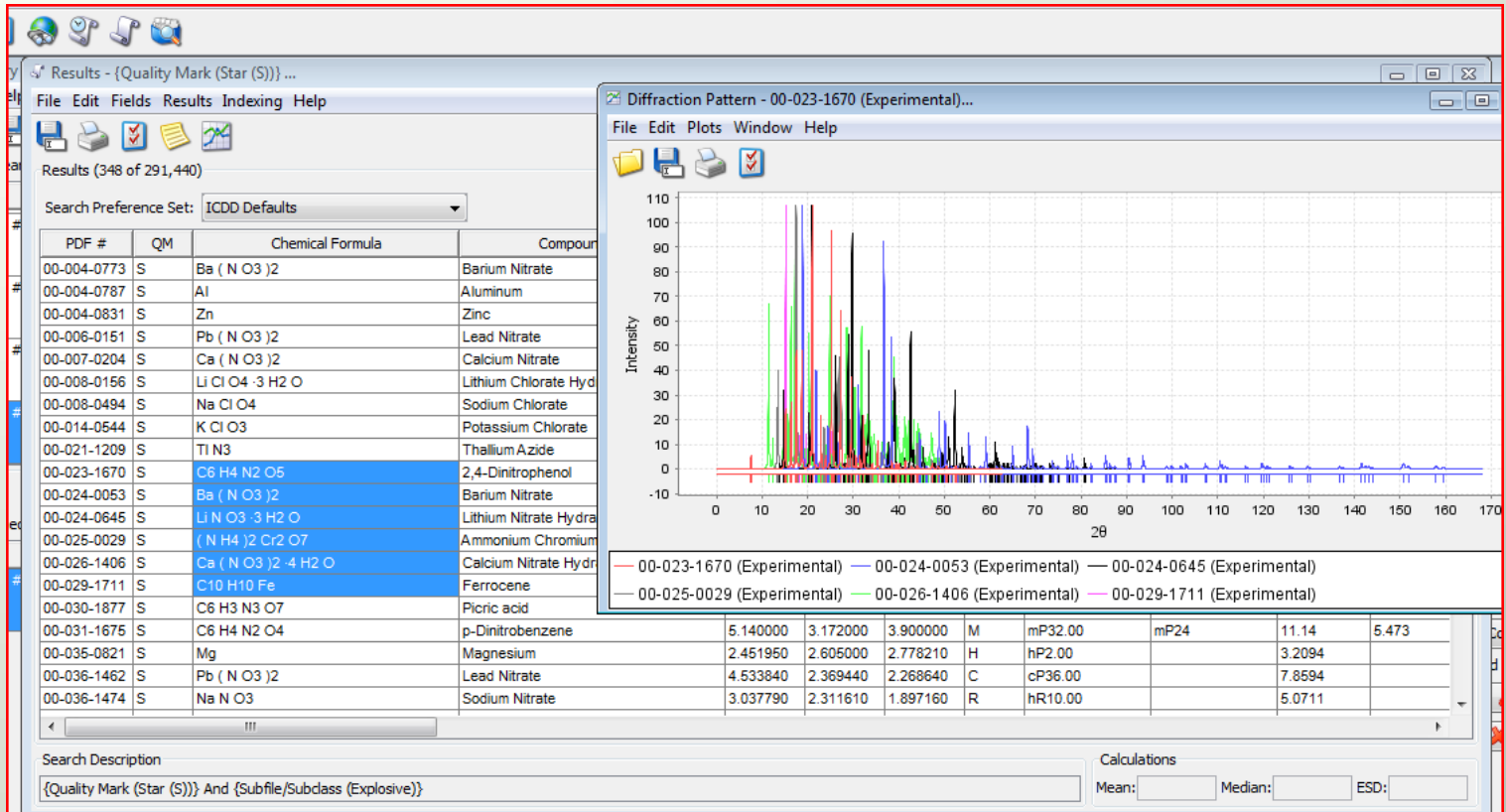
Integral Index – does a point by point comparison of imported diffraction patterns in comparison to dynamically generated experimental and calculated patterns (see tutorial on integral index).

Each pattern is thousands of points, so if large collections of patterns are used in the calculation, then the integral index calculation takes time.

Digital Pattern Calculations

For digital patterns, three algorithms are used depending on the type of data available in the reference material. The most resource-intensive algorithm is used by calculations of patterns from atomic coordinates. Calculation of a single pattern is done in less than a second. Calculation of thousands of patterns takes minutes.

Results Form



The screenshot shows the ICDD software interface. The main window displays search results for 'Quality Mark (Star (S))' with 348 of 291,440 results. The search preference is set to 'ICDD Defaults'. A table of results is shown, with six entries highlighted in blue. To the right, a window titled 'Diffraction Pattern - 00-023-1670 (Experimental)...' displays a plot of Intensity versus 2θ. The plot shows multiple overlapping diffraction patterns for the highlighted entries. A legend at the bottom of the plot identifies the patterns by their PDF numbers and experimental status.

PDF #	QM	Chemical Formula	Compound
00-004-0773	S	Ba (N O3)2	Barium Nitrate
00-004-0787	S	Al	Aluminum
00-004-0831	S	Zn	Zinc
00-006-0151	S	Pb (N O3)2	Lead Nitrate
00-007-0204	S	Ca (N O3)2	Calcium Nitrate
00-008-0156	S	Li Cl O4 · 3 H2 O	Lithium Chlorate Hyd
00-008-0494	S	Na Cl O4	Sodium Chlorate
00-014-0544	S	K Cl O3	Potassium Chlorate
00-021-1209	S	Tl N3	Thallium Azide
00-023-1670	S	C6 H4 N2 O5	2,4-Dinitrophenol
00-024-0053	S	Ba (N O3)2	Barium Nitrate
00-024-0645	S	Li N O3 · 3 H2 O	Lithium Nitrate Hydra
00-025-0029	S	(N H4)2 Cr2 O7	Ammonium Chromium
00-026-1406	S	Ca (N O3)2 · 4 H2 O	Calcium Nitrate Hydr
00-029-1711	S	C10 H10 Fe	Ferrocene
00-030-1877	S	C6 H3 N3 O7	Picric acid
00-031-1675	S	C6 H4 N2 O4	p-Dinitrobenzene
00-035-0821	S	Mg	Magnesium
00-036-1462	S	Pb (N O3)2	Lead Nitrate
00-036-1474	S	Na N O3	Sodium Nitrate

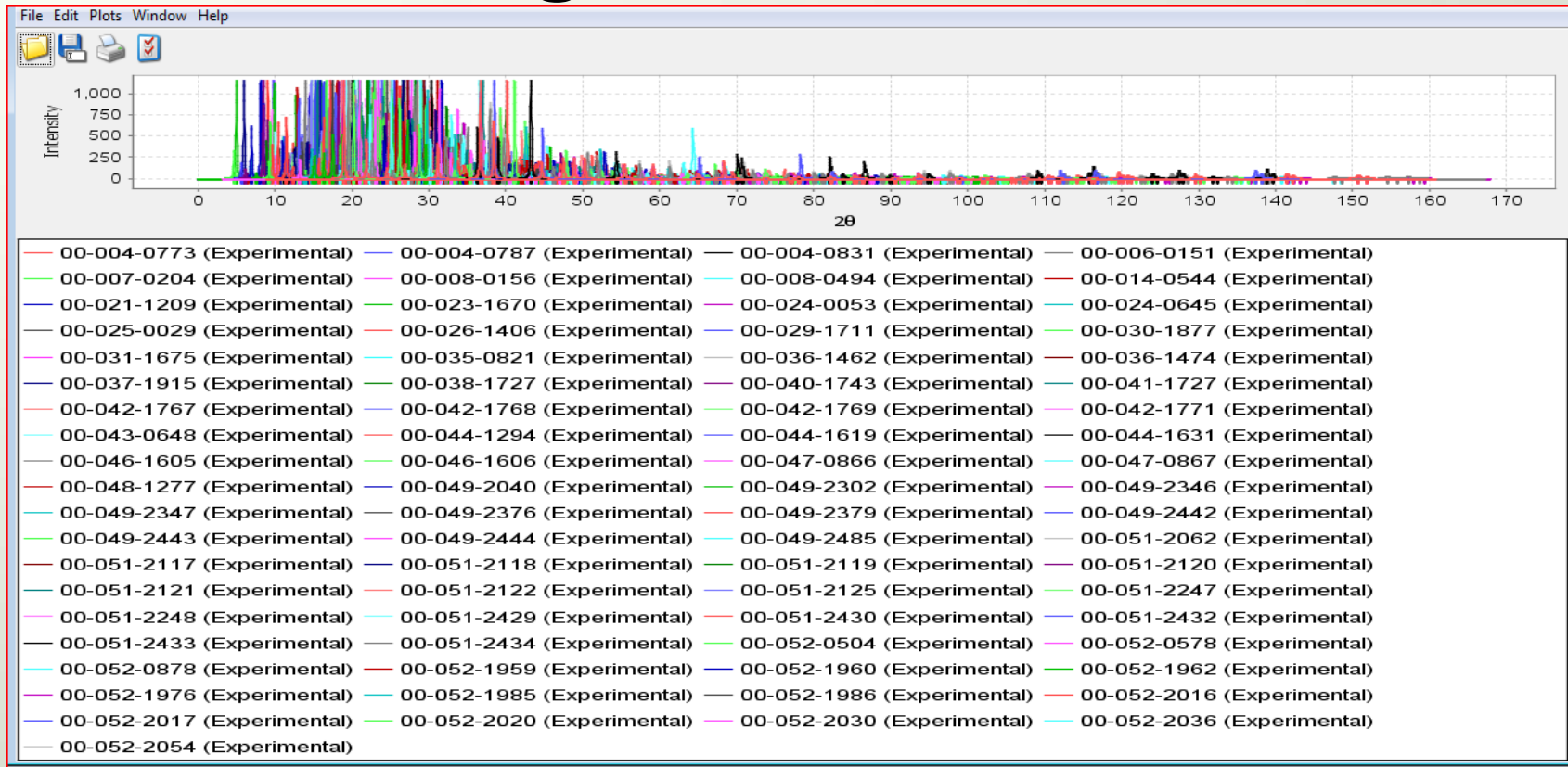
PDF #	QM	Chemical Formula	Compound	5.140000	3.172000	3.900000	M	mP32.00	mP24	11.14	5.473
00-031-1675	S	C6 H4 N2 O4	p-Dinitrobenzene	2.451950	2.605000	2.778210	H	hP2.00		3.2094	
00-036-1462	S	Pb (N O3)2	Lead Nitrate	4.533840	2.369440	2.268640	C	cP36.00		7.8594	
00-036-1474	S	Na N O3	Sodium Nitrate	3.037790	2.311610	1.897160	R	hR10.00		5.0711	

Search Description: {Quality Mark (Star (S))} And {Subfile/Subclass (Explosive)}

Calculations: Mean: Median: ESD:

In this example, six entries were highlighted and a “click” on the right hand button of the mouse produces all six patterns, nearly instantaneously.

Digital Patterns



This calculation of 77 superimposed explosives patterns took several seconds.

The more entries in the calculation, the longer it takes.

Capability Trade Offs

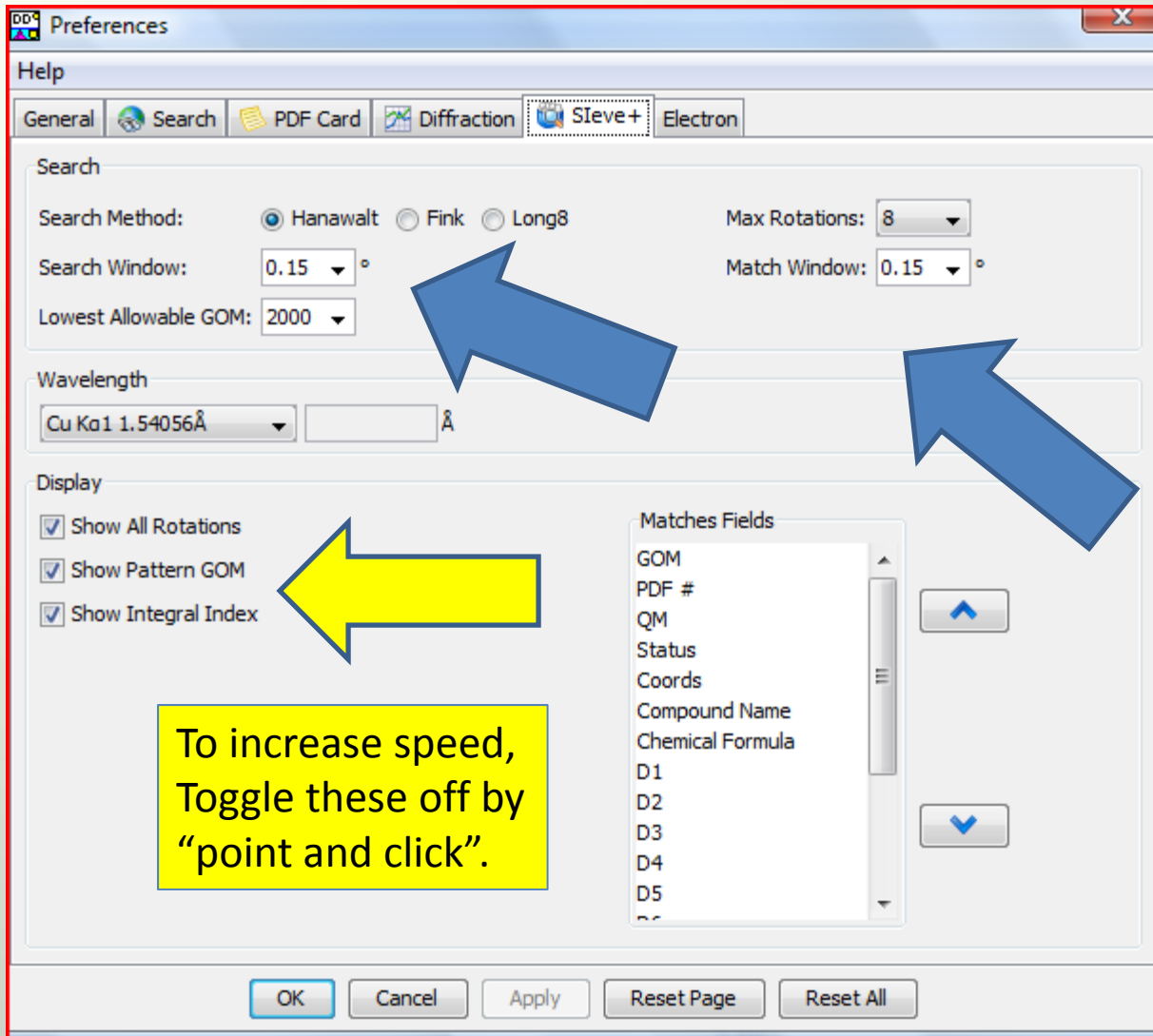
Database searches are frequently used for data mining. Many data mining examples are provided in the tutorials.

A preferred technique in data mining is to use broad search parameters, analyze the results and then apply more restrictions as you find materials of interest. For example, a researcher might analyze all zirconia-containing materials, then focus on yttria or ceria stabilized zirconia and then reduce the candidate list to tetragonally stabilized zirconias, if they were studying cutting tool compositions.

To be effective in the trade-off between capability and speed, one might want to do computationally intensive calculations (i.e. pattern calculation, Integral index) in the later stages of data mining, when the candidate list has been narrowed.

Preference Options in Sieve+

Sieve+



This is the preferences Table for Sieve+.

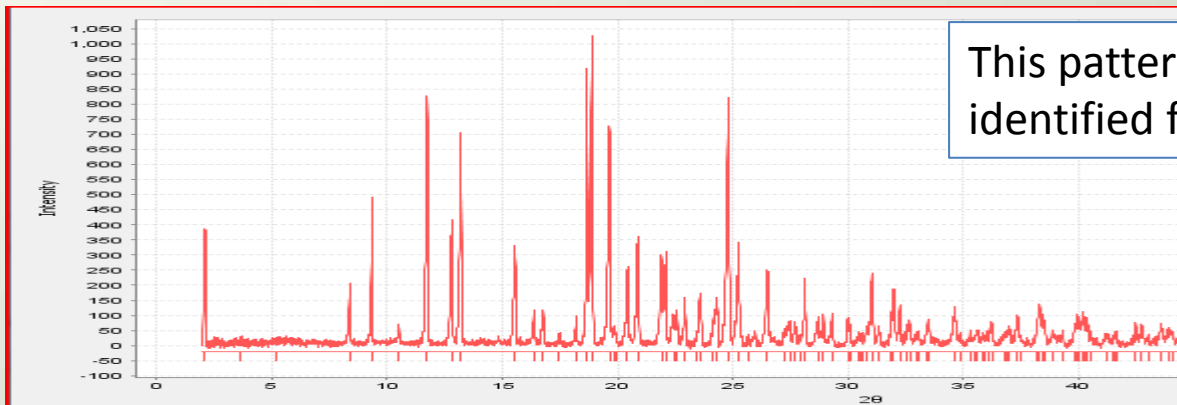
There are several options that can increase or decrease the search speed.

Pattern GOM and Integral Index both involve time-consuming calculations.

To increase speed, Toggle these off by "point and click".

Search and Match Windows and GOM (Goodness of Match)

The wider the search and match windows, the more candidates are reviewed and selected. The more data being compiled for display, the slower the speed.

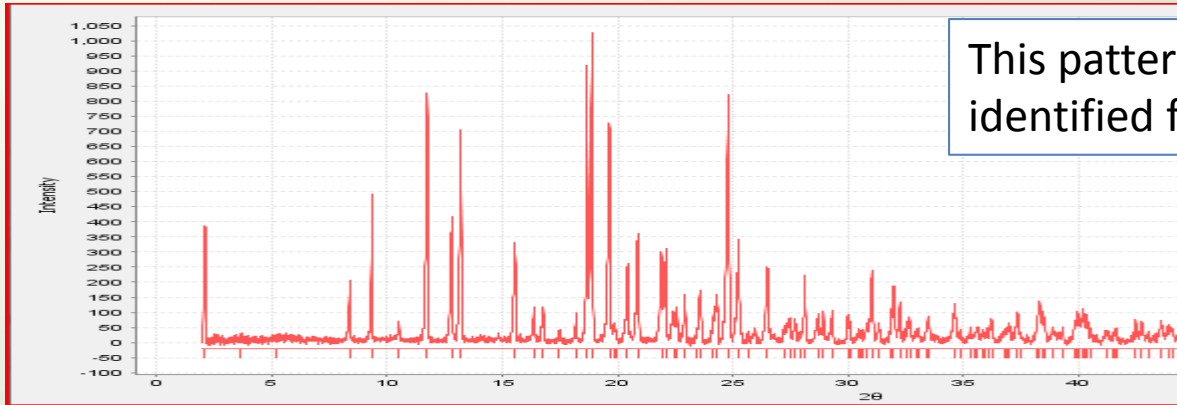


This pattern has 118 peaks identified for match.

<u>Search Window</u>	<u>Match Window</u>	<u>GOM Limit</u>	<u>No of Candidates</u>	<u>Time</u>
0.15	0.15	2000	5,885	5 sec
0.15	0.15	4000	804	< 1 sec
0.06	0.06	4000	10	<<1 sec
0.06	0.06	1000	1,136	< 1sec
0.18	0.18	1000	11,450	5 sec
0.18	0.06	1000	4,176	1 sec

Note: In this example pattern, GOM and integral index calculations were turned OFF.

Integral Index and Pattern GOM



This pattern has 118 peaks identified for match.

<u>Search Window</u>	<u>Match Window</u>	<u>GOM Limit</u>	<u>No of Candidates</u>	<u>Time</u>	<u>Preferences</u>
0.18	0.06	1000	4,176	1 sec	None
0.18	0.06	1000	4,176	5 minutes	Pattern GOM
0.18	0.06	1000	4,176	50 minutes	Integral Index

This is the same example as the previous slide. However, in the second case, the pattern GOM was being calculated for all candidate materials, this took 5 minutes. The integral index was calculated for 4,176 materials and this took 50 minutes.

Capability Trade Offs

Large search and match windows are required when you suspect that you have poor quality data or a high chance of sample displacement errors.

If you have high quality data (standardized data sets), you should be narrowing both these windows.

GOM, Goodness of Match, calculations are based on a scale of 1-8000. Using a high GOM limit (i.e. >4000) means that you will be able to identify major phases, but you are unlikely to identify minor and trace phases where you may only have a few characteristic d-spacings above noise levels. A good strategy is to lower the GOM limit when you are looking for minor phases.

Integral index, while computationally complex, is a basic similarity index that can identify materials independent of the material crystallinity. The search can also be adapted to defined crystallite size ranges. However, if the unknowns are highly crystalline, this calculation can be a significant detriment to speed and productivity, we recommend that you turn it off with highly crystalline data sets.