

# Биг-дата анализ и создание новых материалов

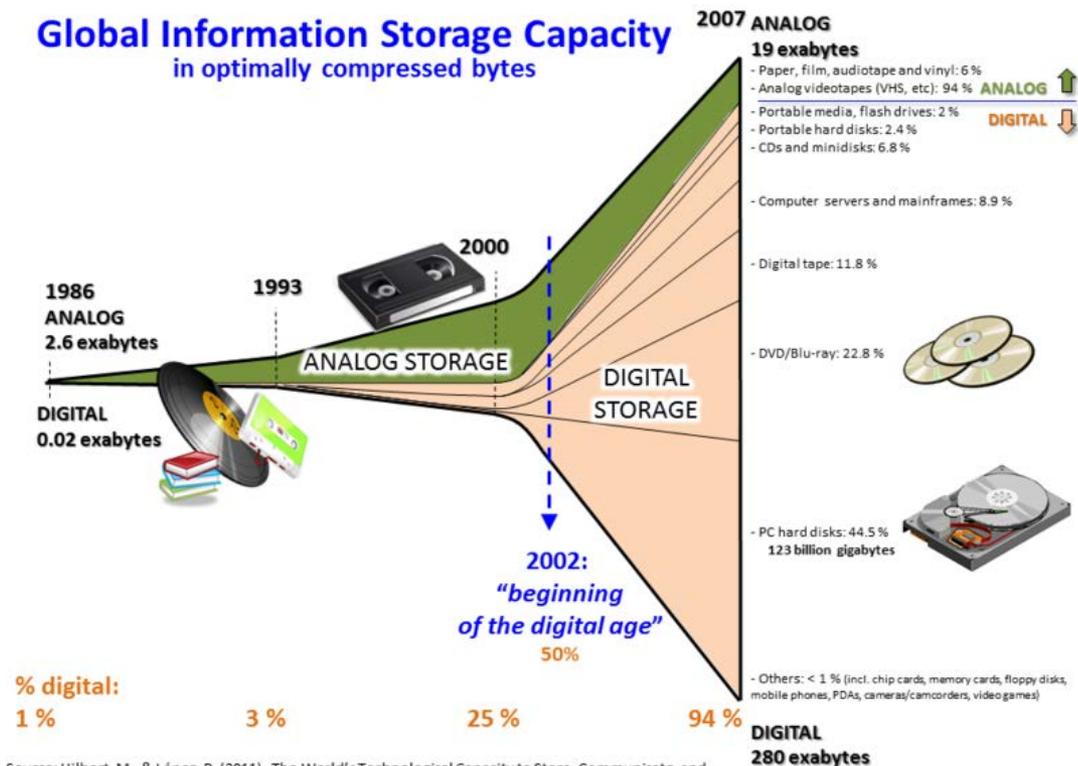
Евгений Александров

Межвузовский научно-исследовательский центр  
по теоретическому материаловедению

Самара, 2018

# Big data определение

- Big data — это инструменты, подходы и методы обработки структурированных и неструктурированных данных для конкретных задач и целей.



Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65. <http://www.martinhilbert.net/WorldInfoCapacity.html>

# Техники и методы анализа

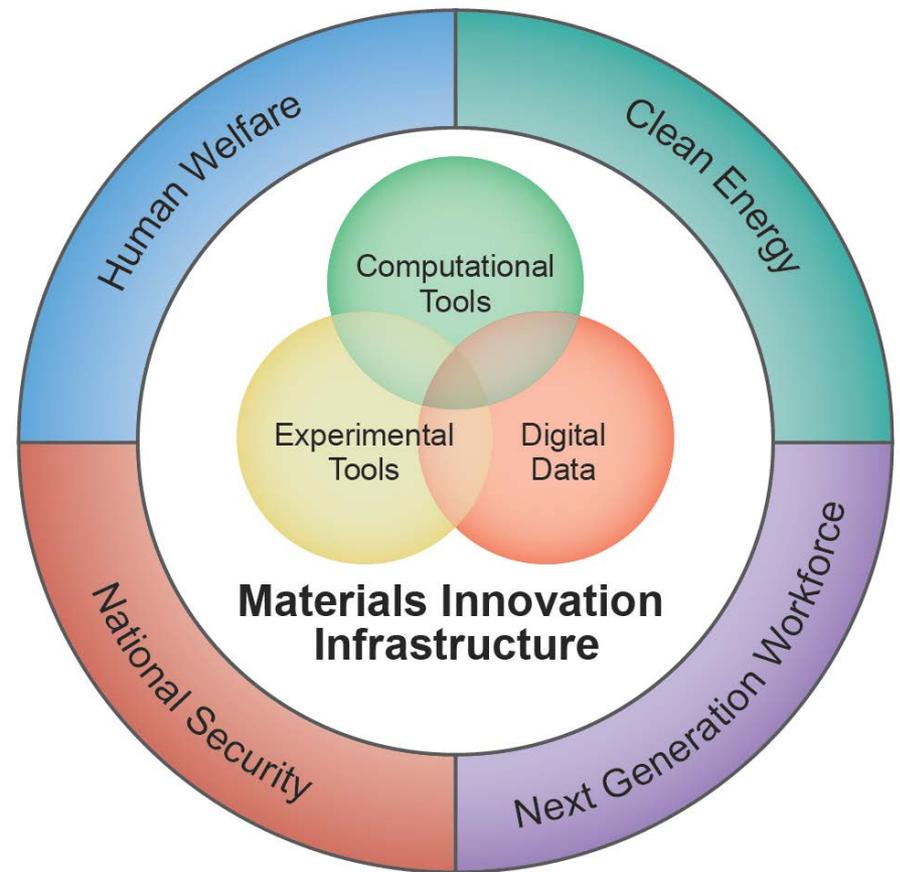
- Data Mining: *обучение ассоциативным правилам, классификация, кластерный анализ, регрессионный анализ;*
- Статистический анализ;
- Визуализация аналитических данных.

# Источники Big-data

- Интернет (соцсети, форумы, блоги, СМИ и другие сайты);
- Корпоративные архивы документов;
- Показания датчиков и приборов, результаты экспериментов.

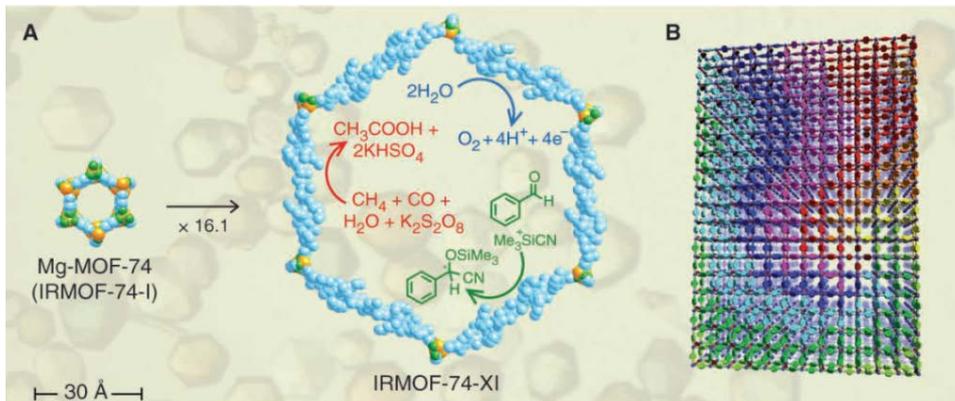
# Materials Genome Initiative for Global Competitiveness

June 2011

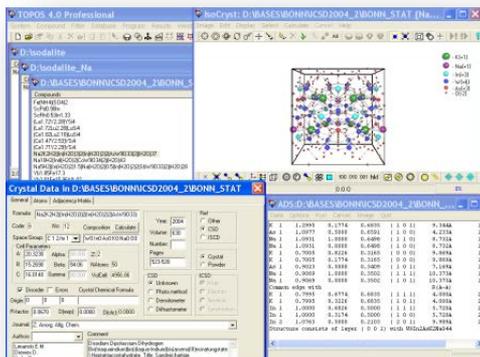


Materials Genome Initiative: A Renaissance of American Manufacturing, Tom Kalil and Cyrus Wadia, <https://www.whitehouse.gov/blog/2011/06/24/materials-genome-initiative-renaissance-american-manufacturing>

# Создание новых материалов



Основная парадигма:  
от микроструктуры к  
макросвойствам



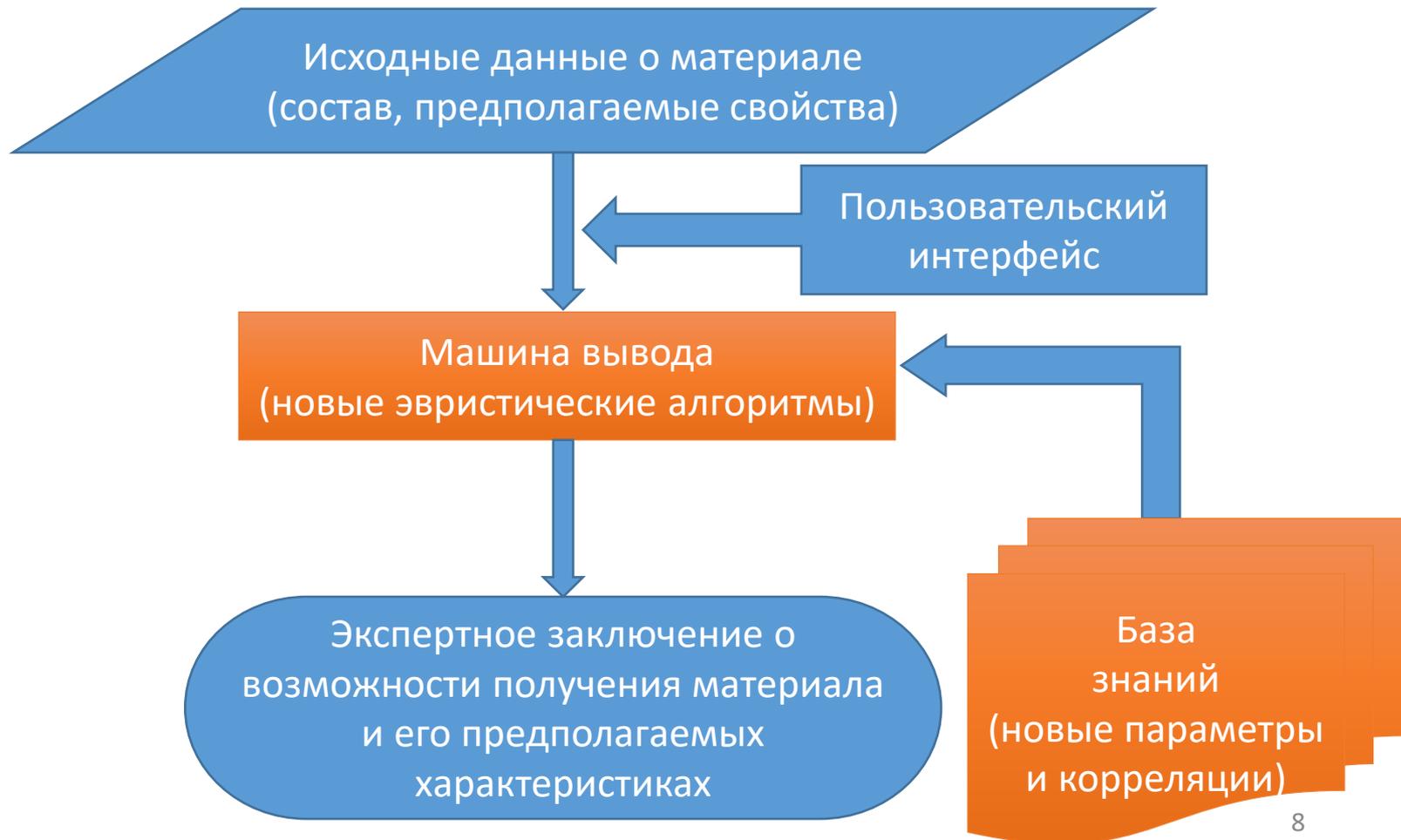
1. Количественное описание и моделирование свойств материалов на атомном и молекулярном уровнях
2. Компьютерный дизайн новых материалов



Лабораторный синтез,  
разработка технологии,  
внедрение



# Экспертные системы в материаловедении

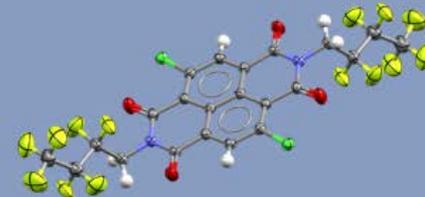


# Данные для поиска корреляций

**Около 900 000 записей**

## The Cambridge Structural Database (CSD)

The world's comprehensive and up-to-date database of crystal structures with over 900,000 curated entries



**Около 870 000 записей**



FIZ Karlsruhe – Leibniz Institute for Information Infrastructure

**Около 154 000 записей**

# Как мы работаем

## Постановка задачи (МГУ, Москва)

Натрий-проводящий катодный материал, содержащий атомы металлов с переменной валентностью, с катионной емкостью не менее 100 мАч/г

## Быстрый поиск аналогов

Более **10000** натрий-содержащих веществ

Более **1000** веществ с атомами металлов с переменной валентностью

Около **300** веществ с катионной емкостью не менее 100 мАч/г

## Быстрый поиск веществ,

которые могут обладать натриевой проводимостью – около **120**

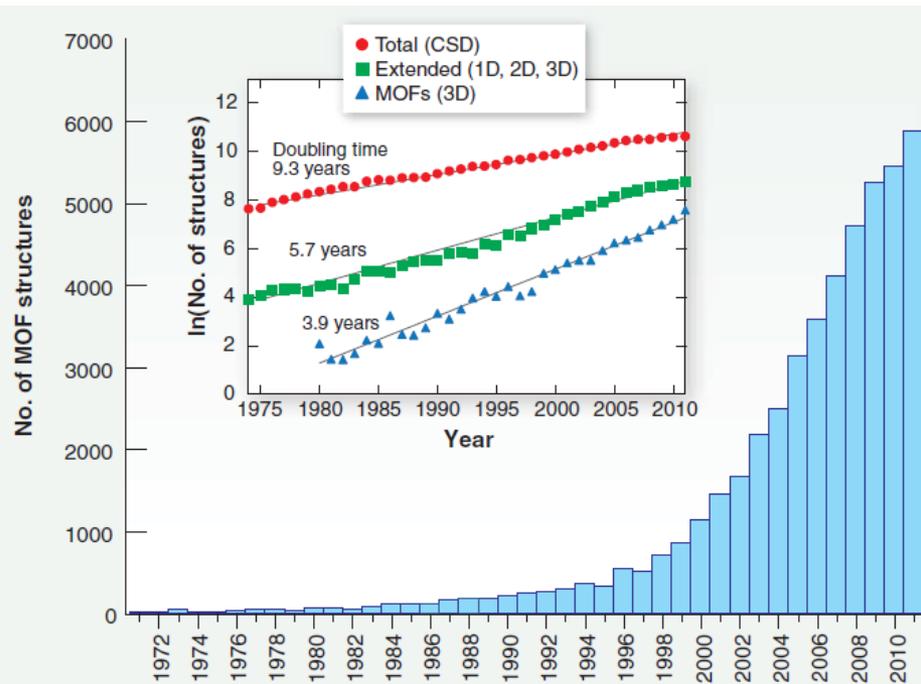
## Детальный анализ полученной выборки

Около **90%** - известные катодные материалы

## Точные расчеты катионной проводимости

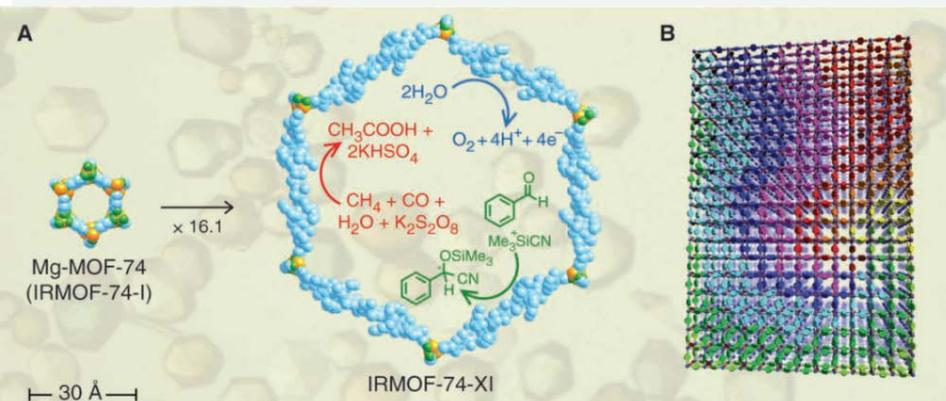
Около **10** веществ представляют собой потенциальные новые катодные материалы

# Координационные полимеры



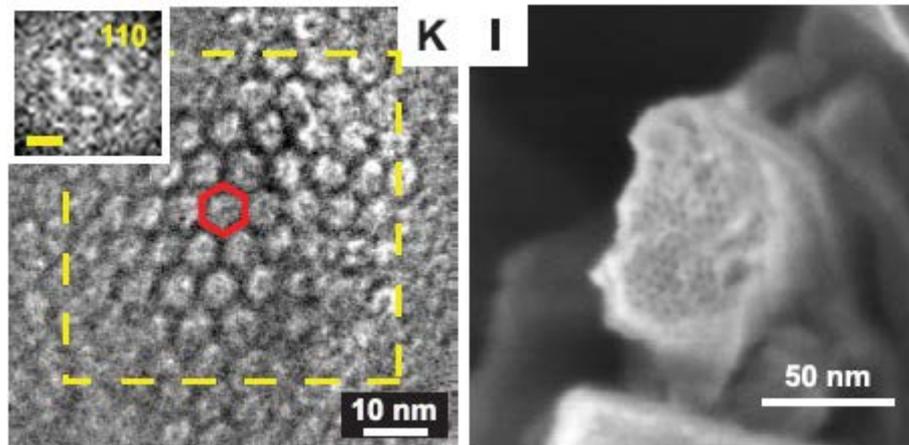
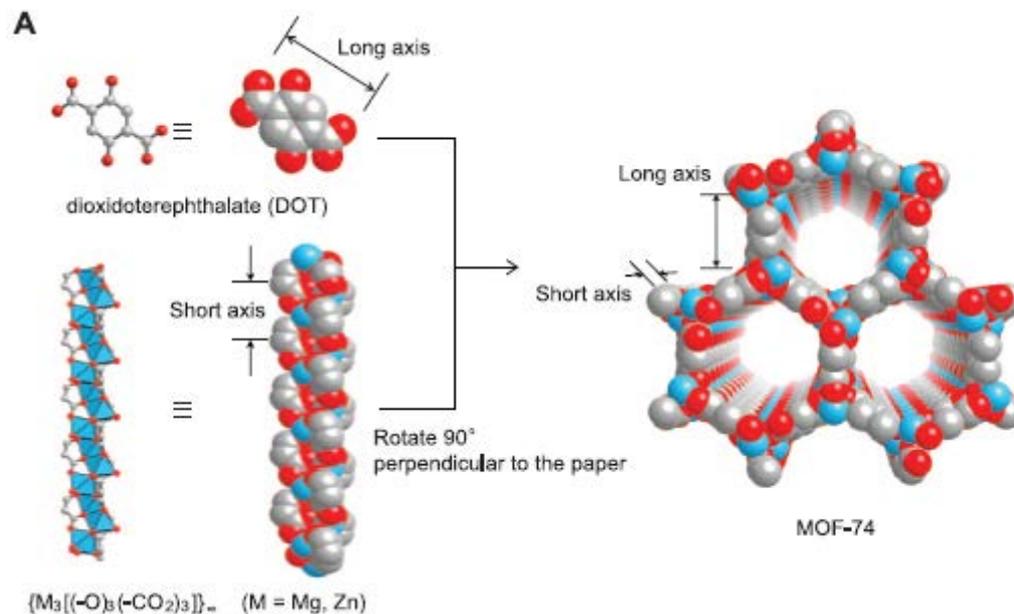
## Полимерные комплексы металлов и органических молекул

- Высокопористые функциональные материалы с наноразмерными монодисперсными порами (газовые ловушки, ионообменники),
- дальний магнитный порядок,
- спиновый кроссовер,
- нелинейная оптическая активность
- хиральность,
- реакционная и каталитическая активность,
- люминесценция и т.д.



H. Furukawa K. E. Cordova, M. O’Keeffe, O.M. Yaghi The Chemistry and Applications of Metal-Organic Frameworks. Science. 341. 6149. 974-976.

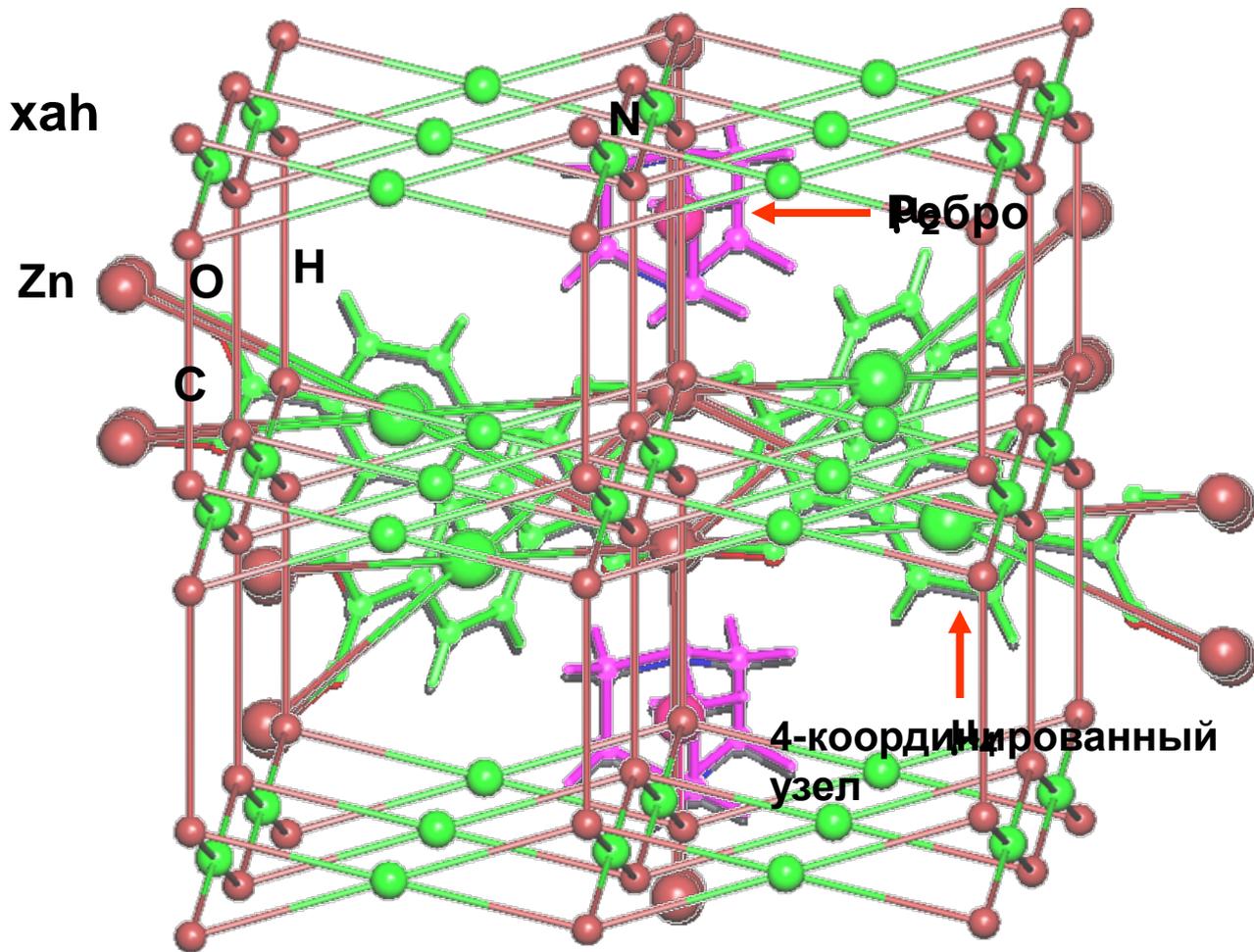
# Пористые координационные полимеры



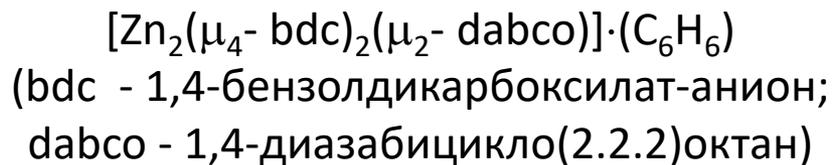
# Геометрико-топологические методы моделирования пористых материалов

- Топологические мотивы связывания строительных единиц
- Топология переплетения координационных полимеров
- Топология системы пустот и каналов в кристаллической структуре

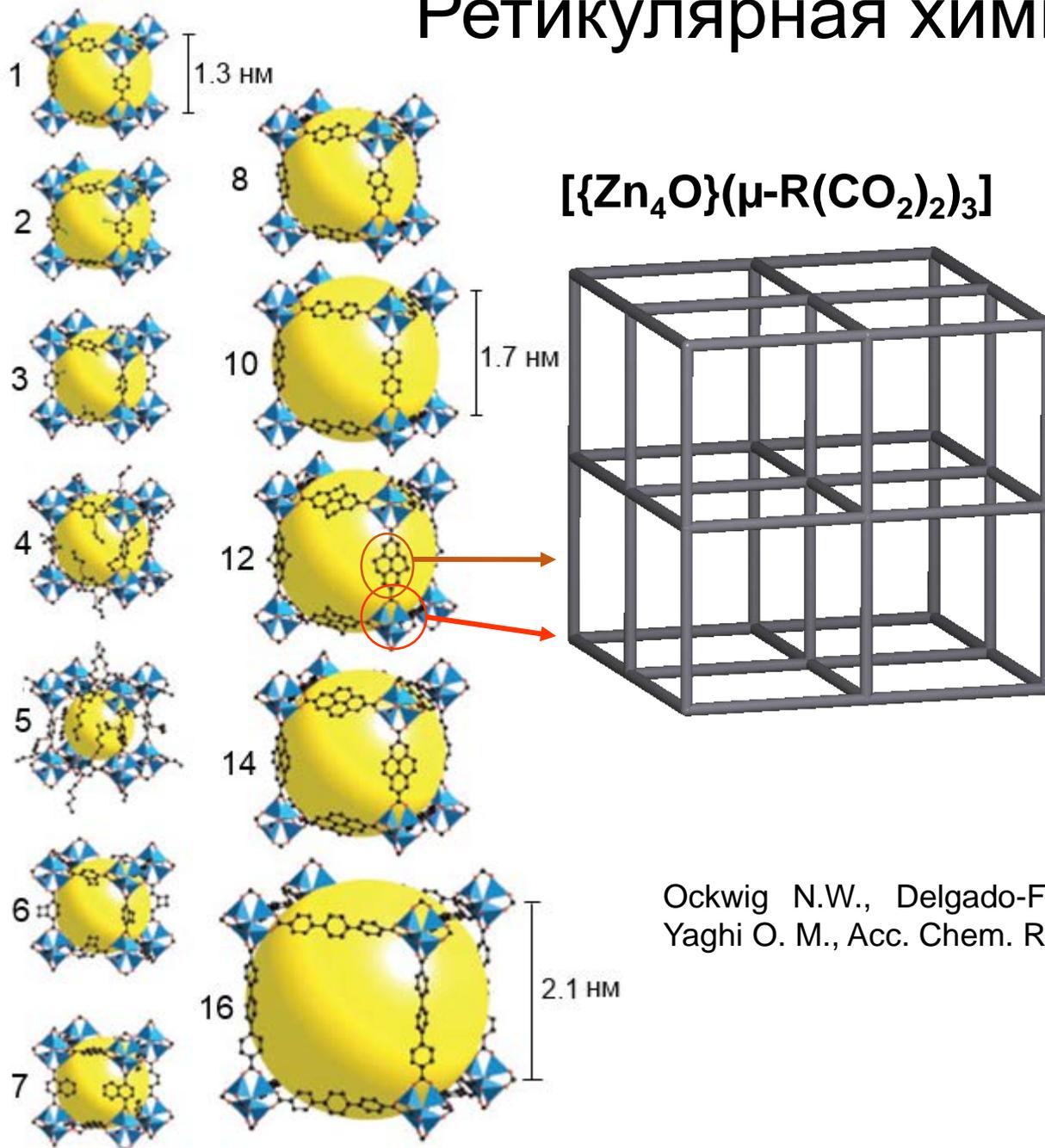
# Топологические мотивы связывания строительных единиц



0-координированные узлы – внекаркасные частицы  
1-координированные узлы – концевые лиганды  
2-координированные узлы – мостиковые  $\mu_2$  лиганды и 2-координированные атомы металла  
n-координированные узлы – мостиковые  $\mu_n$  лиганды и n-координированные атомы металла ( $n > 2$ )



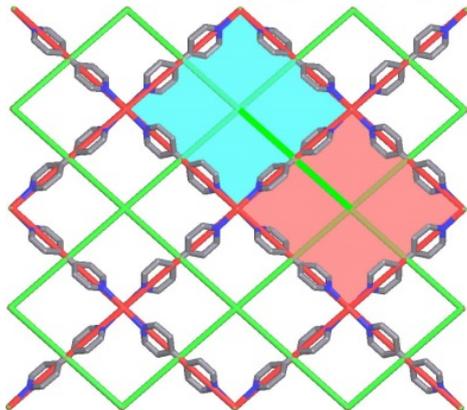
# Ретикулярная химия



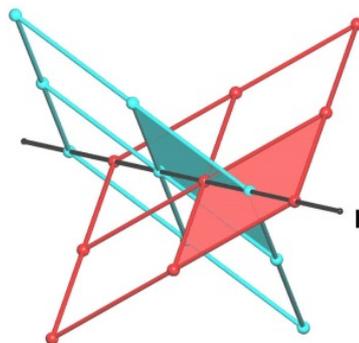
Ockwig N.W., Delgado-Friedrichs O., O'Keeffe M., Yaghi O. M., Acc. Chem. Res. 2005. 38(3). 176–182.

# Переплетения

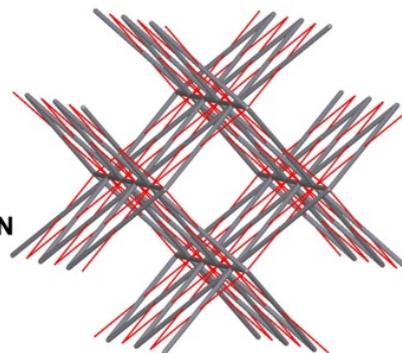
JEZRUB  $[\text{Zn}(\text{H}_2\text{O})_2(4,4'\text{-bipy})_2]^{2+}$



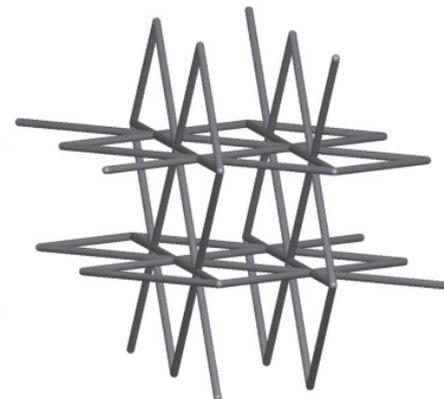
underlying net (red) **sq1** +  
fused ring net (green) **FRN sq1**



inclined catenation of two **sq1** +  
*Hopf ring net (HRN) 2-c linear chain*

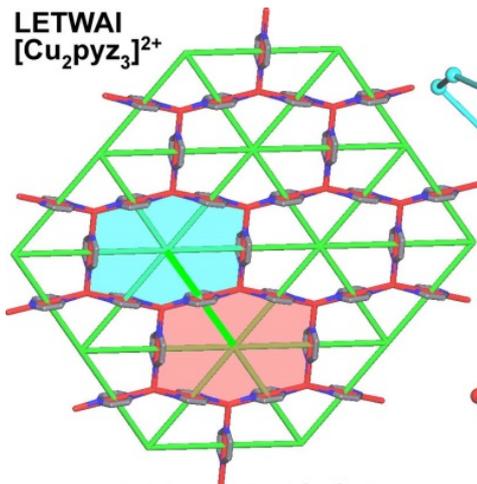


inclined catenation of **sq1** (red) +  
6-c extended ring net (grey) **ERN**

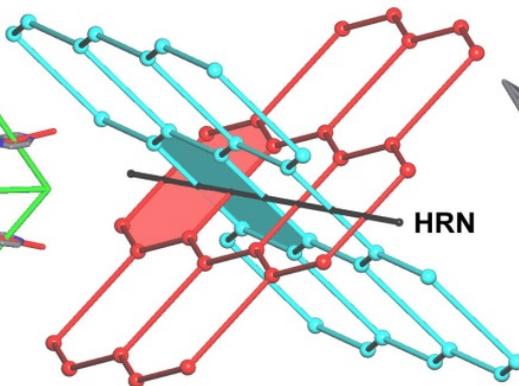


single 6-c extended ring net  
 $\text{ERN} = \text{FRN} + \text{HRN} = (4^9.6^6)\text{-}6\text{T}13$

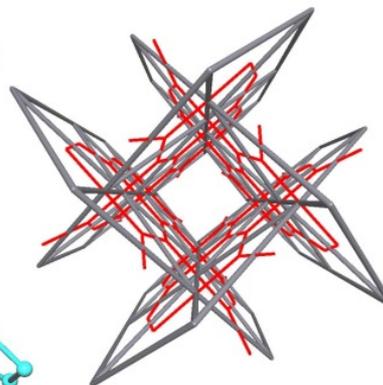
LETWAI  
 $[\text{Cu}_2\text{pyz}_3]^{2+}$



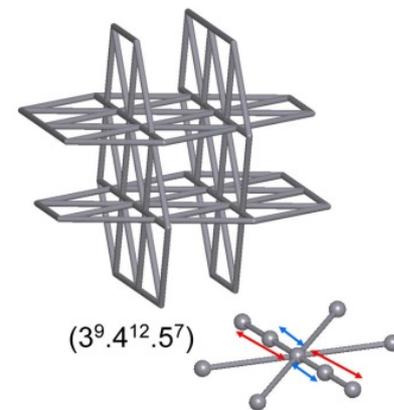
underlying net (red) **hcb** +  
fused ring net (green) **FRN hxl**



inclined catenation of two **hcb** +  
*Hopf ring net (HRN) 2-c linear chain*



inclined catenation of **hcb** (red) +  
8-c extended ring net (grey) **ERN**

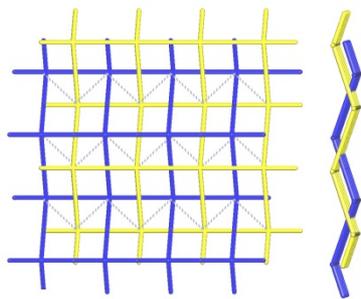
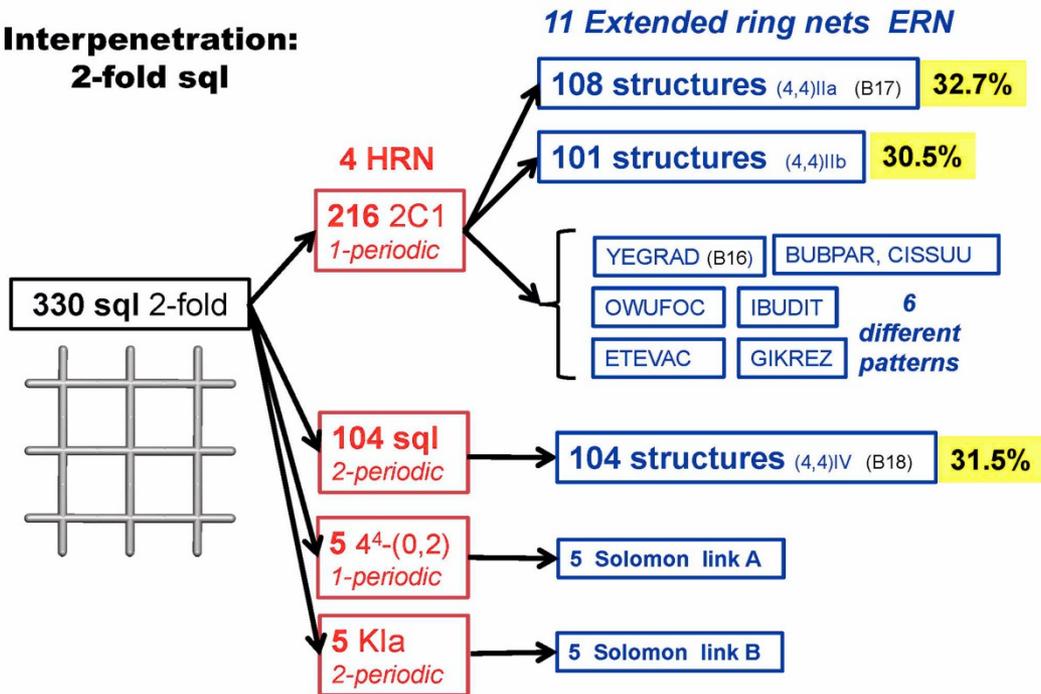


$(3^9.4^{12}.5^7)$

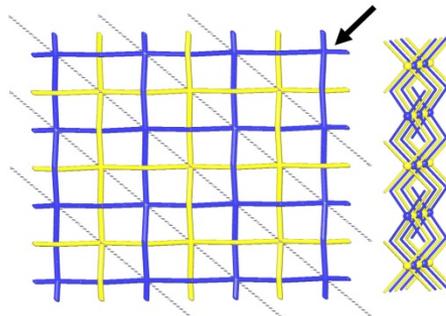
single 8-c extended ring net  
with some coincident edges

# Типы переплетения

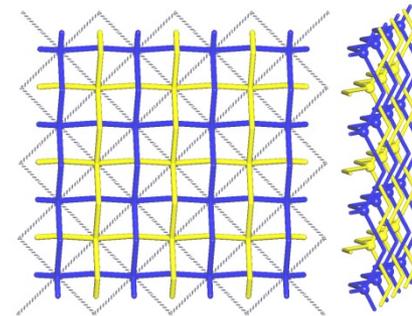
**Interpenetration:  
2-fold sq1**



**108 structures** ERN (4,4)IIa B17

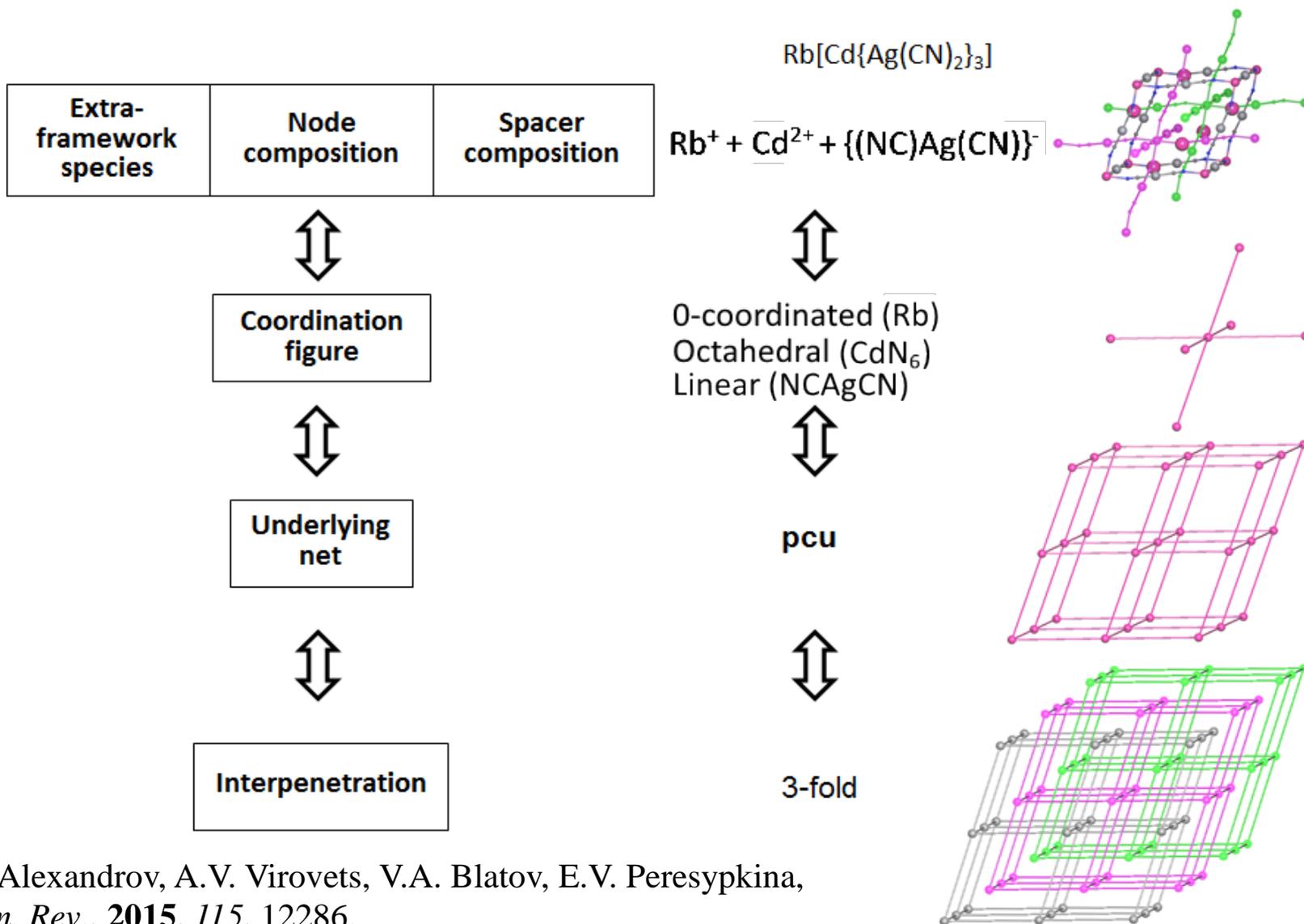


**101 structures** ERN (4,4)IIb

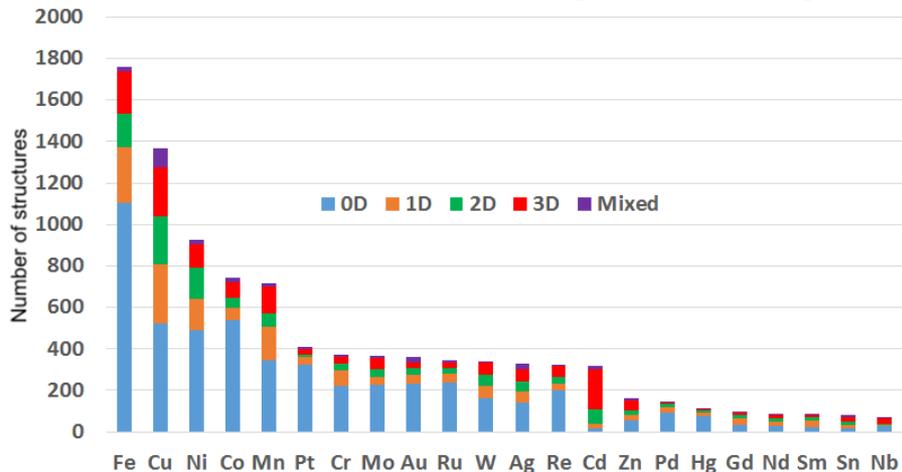


**104 structures** ERN (4,4)IV B18

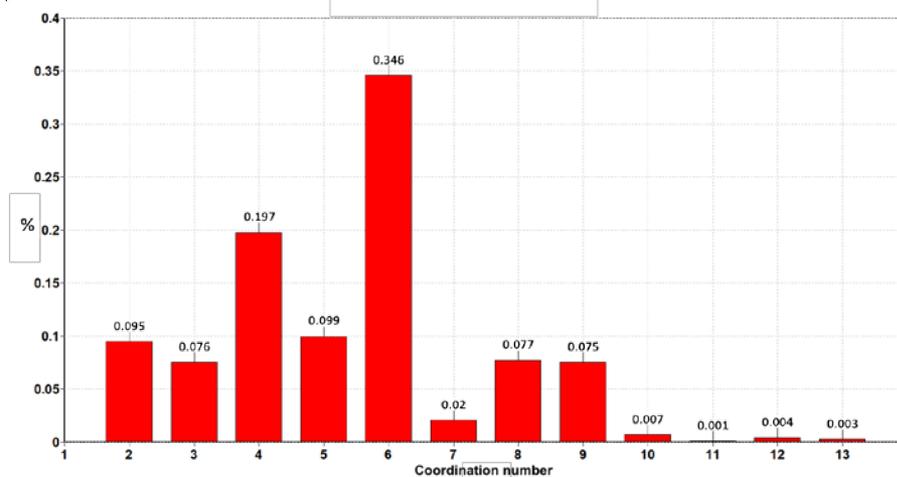
# Схема многоуровневого топологического анализа



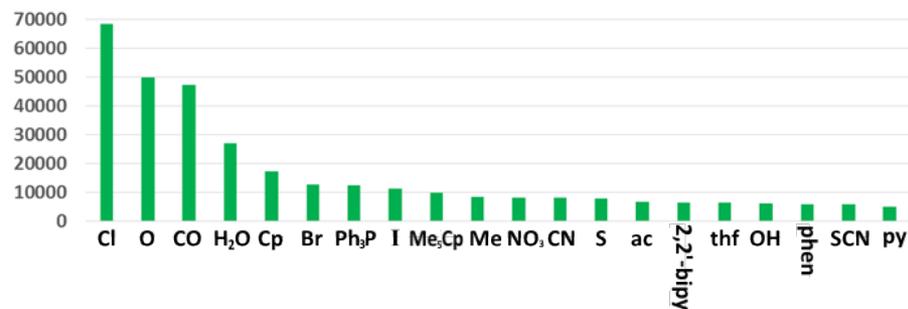
# Построение статистических распределений



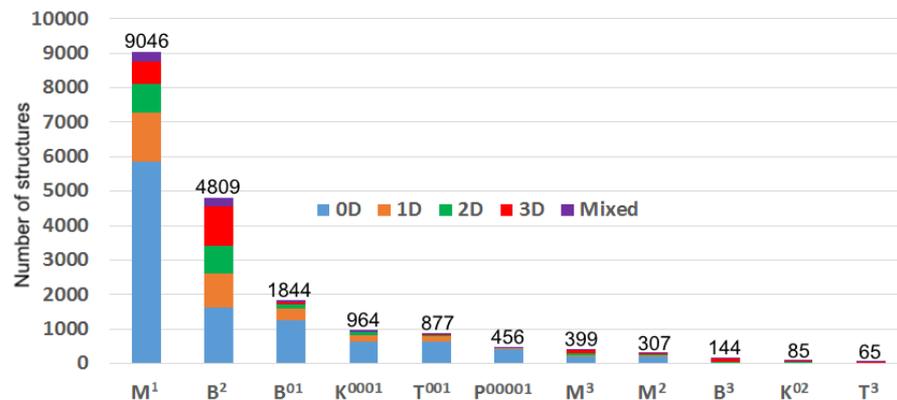
Центральные атомы



Координационное число

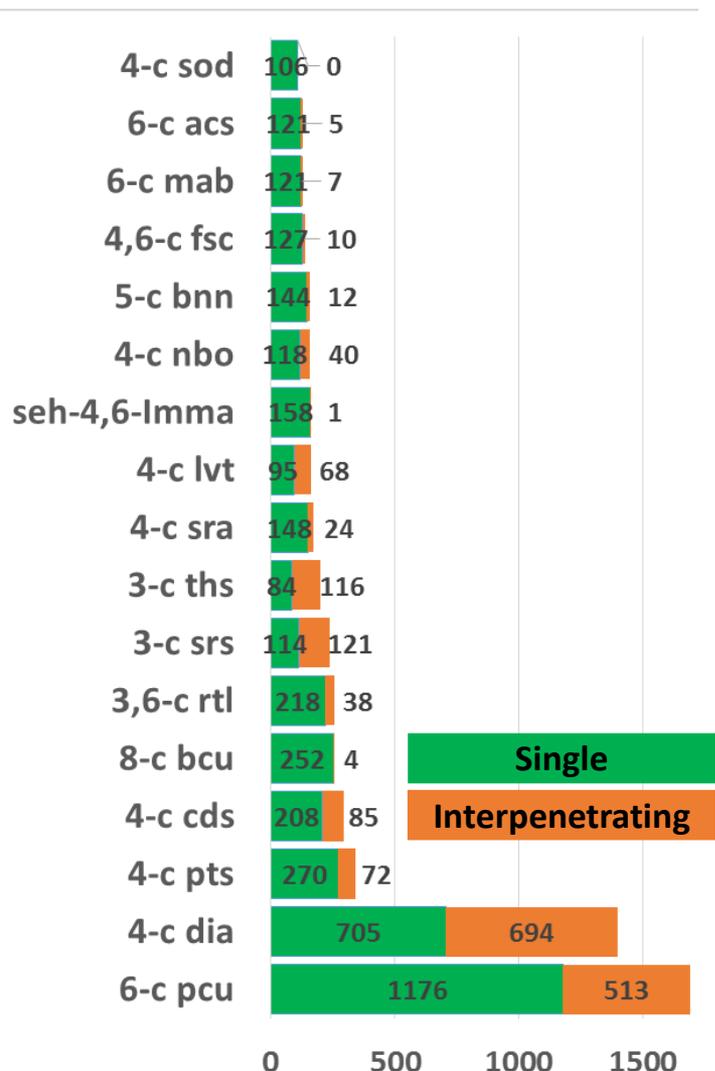


Лиганды



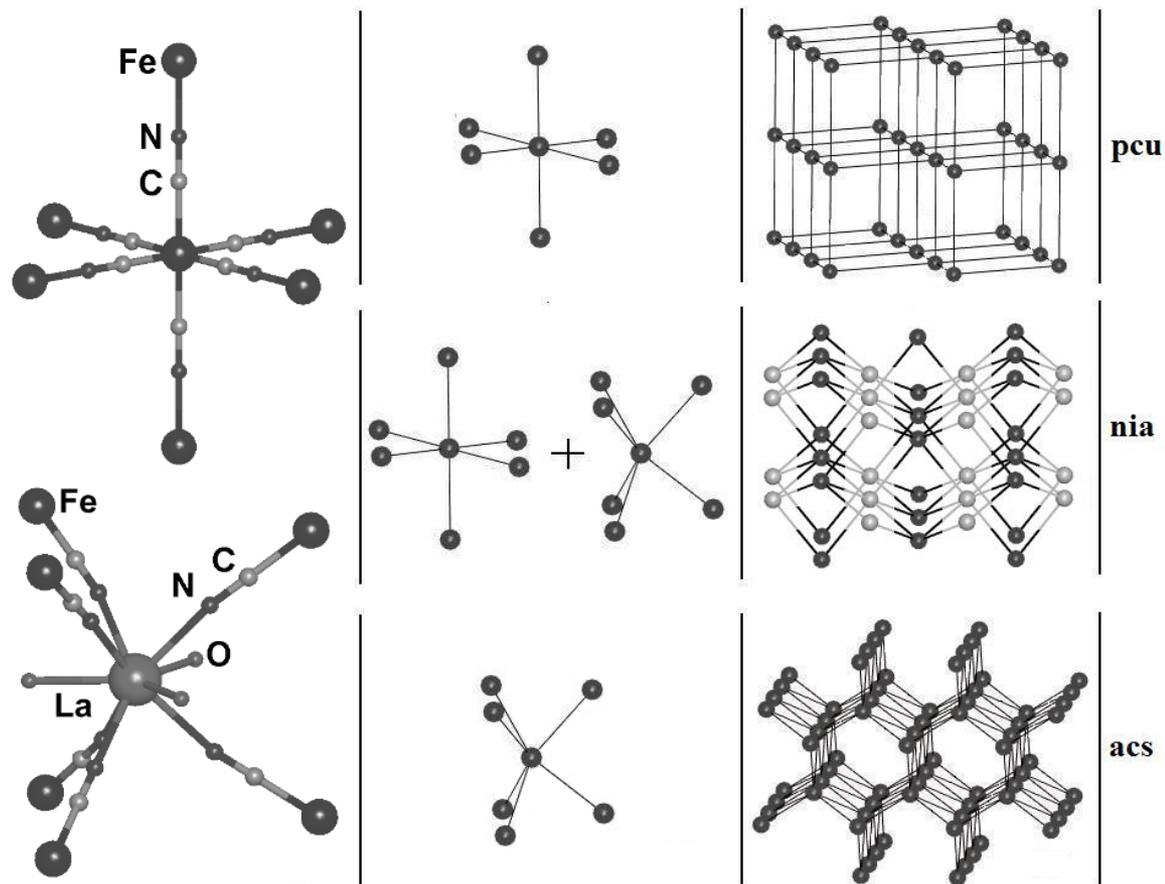
Координация лигандов

# 18470 структур – 4813 топологических ТИПОВ



<i>pqrs</i>	Описание топологического типа
1121	Sodalite, semiregular, simple uniform tiling
1122	Spherical packing (SP) 6/4/h2, semiregular net
1231	Pillared <b>sql</b> layers
2222	Theoretical <i>sqc11</i>
1221	BN, uniform tiling, SP 5/4/h5, <i>sqc12</i> , invariant rod packing
<b>1111</b>	NbO, regular, SP 4/6/c2, <i>sqc35</i> , invariant rod packing
2541	Theoretical, rod-packing <b>sra</b> or <b>pcu</b>
1121	Semiregular net, SP 4/4/t1, <i>sqc176</i>
1331	SrAl <sub>2</sub> , zeolite ABW, quasisimple tiling, SP 4/4/o1, rod-packing
1211	ThSi <sub>2</sub> , minimal uniform, SP 3/10/t4
<b>1111</b>	SrSi <sub>2</sub> , minimal regular uniform self-dual chiral, SP 3/10/c1
2232	Rutile, rod packing <b>pcu</b>
<b>1111</b>	BCC, minimal regular self-dual net, SP 8/4/c1, <i>sqc3</i>
1221	CdSO <sub>4</sub> , minimal uniform self-dual, SP 4/6/t4, <i>sqc5</i> , invariant RP
2132	PtS, theoretical <i>sqc183</i> , rod-packing
<b>1111</b>	Diamond, minimal regular uniform self-dual, quasisimple tiling, SP 4/6/c1, <i>sqc6</i>
<b>1111</b>	$\alpha$ -Po, minimal regular uniform self-dual, uniform tiling, SP 6/4/c1, <i>sqc1</i> , invariant rod packing

# Корреляции «состав-координация-топология»



Александров Е. В., Пономарева А. С., Блатов В. А. Коорд. химия. 2011. 37(2). 83-95.

# Ассоциативные правила

## КООРДИНАЦИЯ МЕТАЛЛА

IF ( $Me='Mn'$ ) AND ( $OxS=2$ ) AND ( $Cn=6$ ) THEN  $CCp='N6'$  with  $P=71.9\%$

IF ( $Me='Mn'$ ) AND ( $OxS=3$ ) AND ( $Cn=6$ ) THEN  $CCp='C6'$  with  $P=90.0\%$

## КООРДИНАЦИОННАЯ ФОРМУЛА

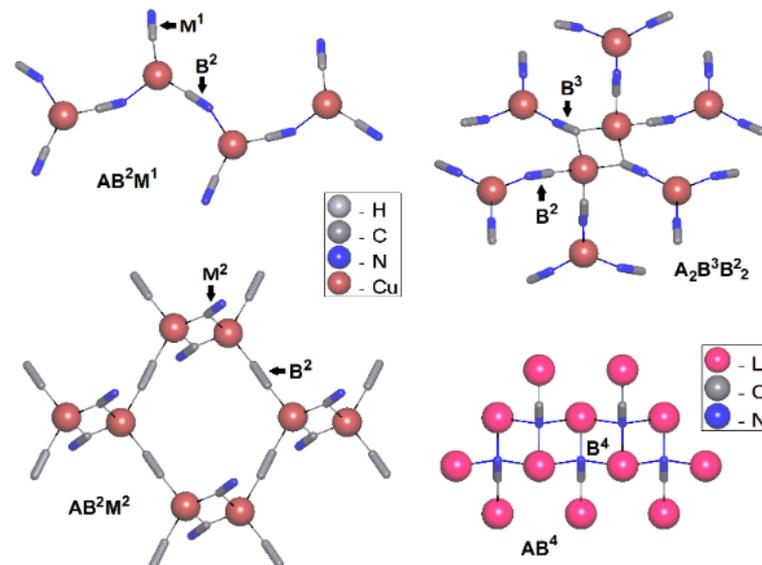
Coordination modes [Serezhkin, V. N. et al. *Acta Cryst. B* 2009, 65, 45]:

$M^1$  (561, 63.2 %),  $B^2$  (303, 34.2 %),  $B^3$  (21, 2.4 %)

Coordination formulas:

$AM^1_6$  (206, 25.4%),  $AM^1_4$  (183, 22.6%),  $AM^1_2$  (98, 12.1%),

$AB^2_3$  (87, 10.7%),  $AB^2_2$  (51, 6.3 %)



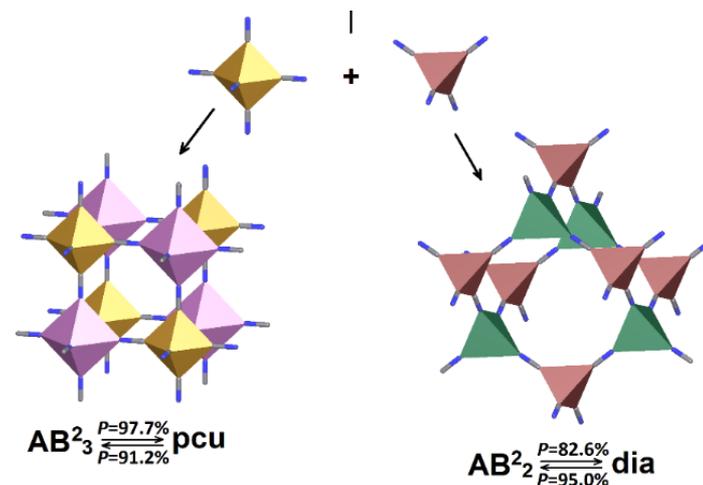
## ТОПОЛОГИЯ СЕТКИ

IF  $CF='AB^2_3'$  THEN  $UT='pcu'$  with  $P=97.7\%$ ;

IF  $UT='pcu'$  THEN  $CF='AB^2_3'$  with  $P=91.2\%$ ;

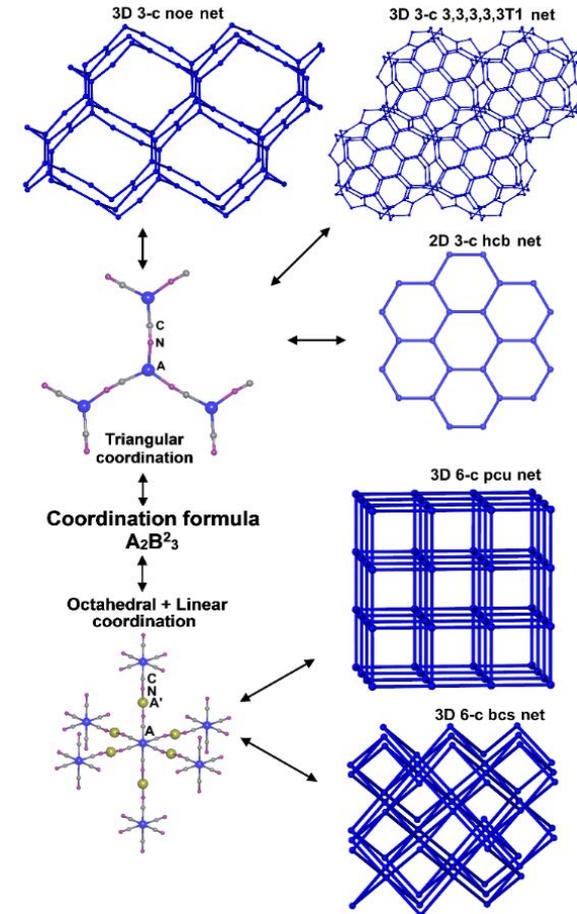
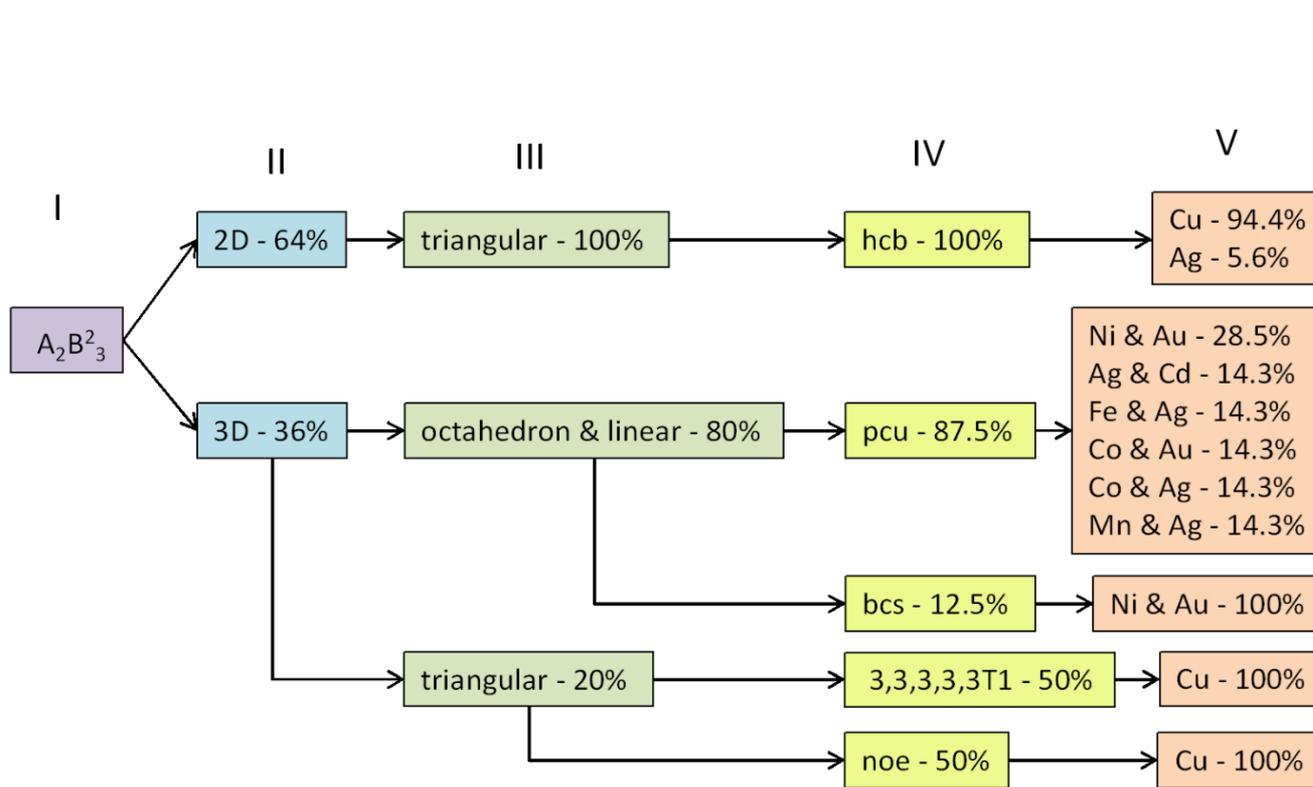
IF  $CF='AB^2_2'$  THEN  $UT='dia'$  with  $P=82.6\%$ ;

IF  $UT='dia'$  THEN  $CF='AB^2_2'$  with  $P=95.0\%$ .



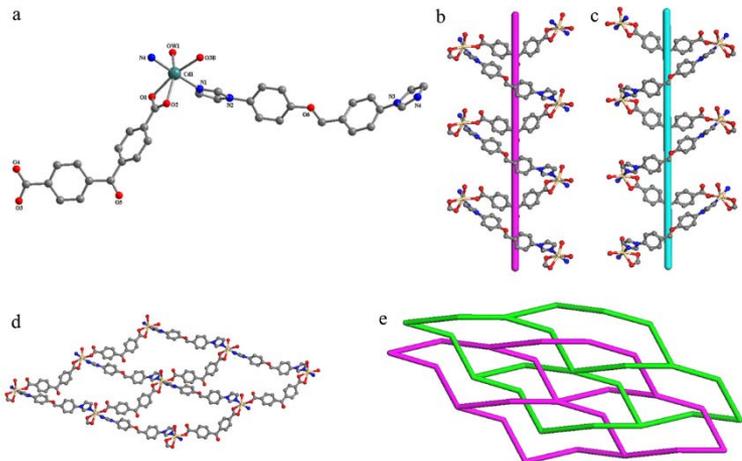
Alexandrov E.V., Shevchenko A.P., Asiri A.A., Blatov V.A.  
*CrystEngComm*, 2015, 17, 2913.

# Дерево решений

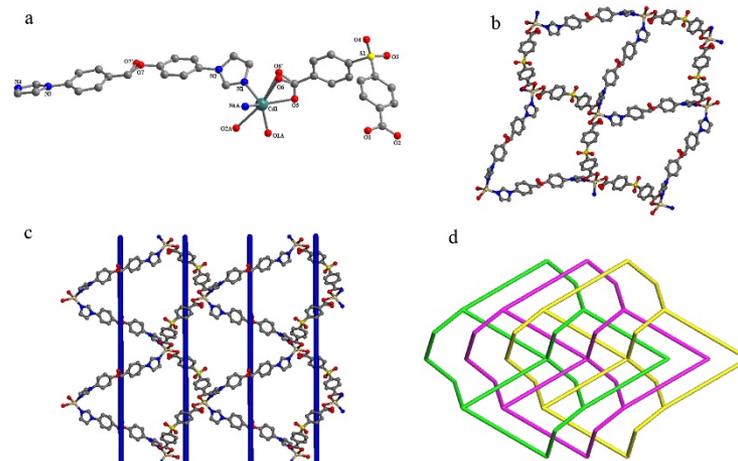


Alexandrov E.V., Shevchenko A.P., Asiri A.A., Blatov V.A.  
*CrystEngComm*, **2015**, *17*, 2913.

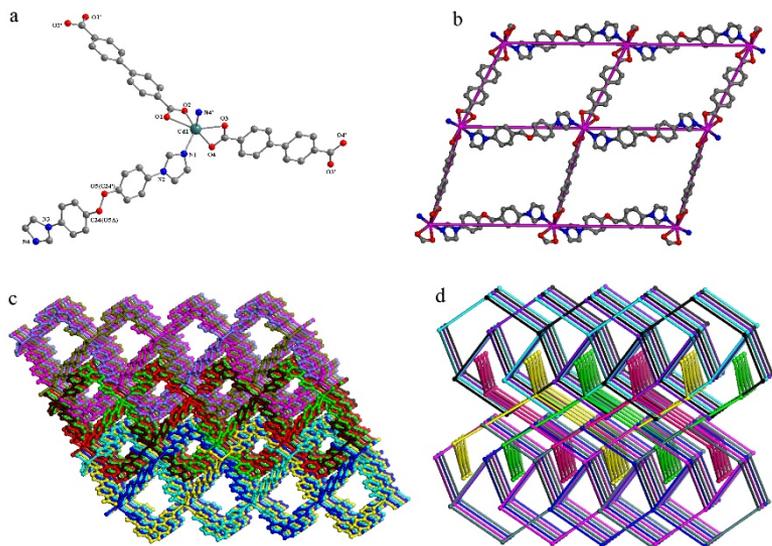
# Примеры синтезированных координационных полимеров



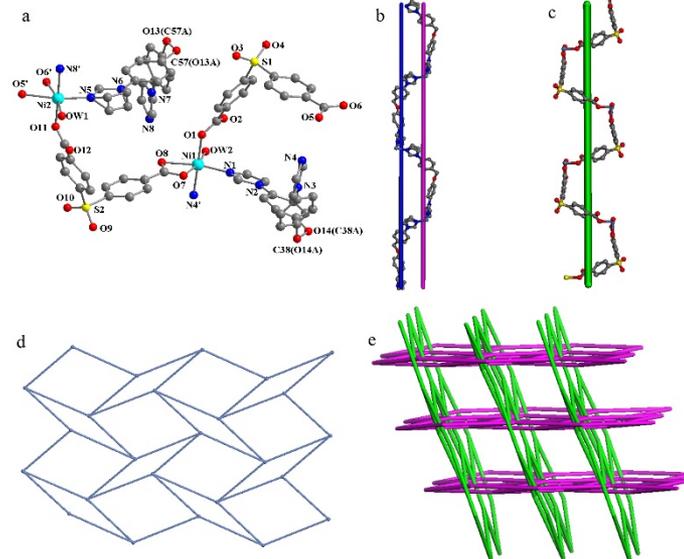
[Cd(cba)(ibpi)] (101 пример)



[Cd(sdc)(ibpi)]·H<sub>2</sub>O (12 примеров)



[Cd<sub>2</sub>(bpdc)<sub>2</sub>(ibpi)<sub>2</sub>] (1 пример)



[Ni<sub>2</sub>(sdc)<sub>2</sub>(ibpi)<sub>2</sub>(H<sub>2</sub>O)<sub>2</sub>]·H<sub>2</sub>O

# Цеолиты

rcsr.fos.su.se/IZA-SC/ft.php

Database of Zeolite Structures

IZA-SC Framework Type References Powder Patterns Building Schemes Disordered Structures Other Links

Advanced Search Credits Help

## Zeolite Framework Types

### Search for a Framework Type Code (FTC)

(Two or more letters/numbers; also search for a material that contains this string)

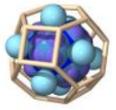
or select an FTC from the table:

ABW	ACO	AEI	AEL	AEN	AET	AFG	AFI	AFN	AFO	AFR	AFS
AFT	AFV	AFX	AFY	AHT	ANA	APC	APD	AST	ASV	ATN	ATO
ATS	ATT	ATV	AVL	AWO	AWW	BCT	*BEA	BEC	BIK	BOF	BOG
BOZ	BPH	BRE	BSV	CAN	CAS	CDO	CFI	CGF	CGS	CHA	-CHI
-CLO	CON	CSV	CZP	DAC	DDR	DFO	DFT	DOH	DON	EAB	EDI
EEL	EMT	EON	EPI	ERI	ESV	ETL	ETR	EUO	*EWT	EZT	FAR
FAU	FER	FRA	GIS	GIU	GME	GON	GOO	HEU	IFO	IFR	-IFU
IFW	IFY	IHW	IMF	IRN	IRR	-IRY	ISV	ITE	ITG	ITH	*-ITN
ITR	ITT	-ITV	ITW	IWR	IWS	IWW	IWW	JBW	JNT	JOZ	JRY
JSN	JSR	JST	JSW	KFI	LAU	LEV	LIO	-LIT	LOS	LOV	LTA
LTF	LTJ	LTL	LTN	MAR	MAZ	MEI	MEL	MEP	MER	MFI	MFS
MON	MOR	MOZ	*MRE	MSE	MSO	MTF	MTN	MTT	MTW	MVY	MWF
MWW	NAB	NAT	NES	NON	NPO	NPT	NSI	OBW	OFF	OKO	OSI
OSO	OWE	-PAR	PAU	PCR	PHI	PON	POS	PSI	PUN	RHO	-RON
RRO	RSN	RTE	RTH	RUT	RWR	RWY	SAF	SAO	SAS	SAT	SAV
SBE	SBN	SBS	SBT	SEW	SFE	SFF	SFG	SFH	SFN	SFO	SFS
*SFV	SFW	SGT	SIV	SOD	SOF	SOS	SSF	*-SSO	SSY	STF	STI
*STO	STT	STW	-SVR	SVV	SZR	TER	THO	TOL	TON	TSC	TUN
UEI	UFI	UOS	UOV	UOZ	USI	UTL	UWY	VET	VFI	VNI	VSV
WEI	-WEN	YUG	ZON								

Синтезированные:  
232 типа каркасов

Яндекс www.hypotheticalzeolites.net

HOME DATABASES TOOLS REFERENCES LINKS STATISTICS ABOUT

 Atlas of Prospective Zeolite Structures

Martin D. Foster and Michael M.J. Treacy

Welcome to the Atlas of prospective zeolite frameworks.

- The website was last updated May 5th, 2010.
- There are currently 5,389,407 structures in the Bronze database.
- There are currently 2,175,685 confirmed Bronze topologies that are unique.
- There are currently 97 of the IZA frameworks in the Bronze database.

If you would like to reference this website, we suggest the following:

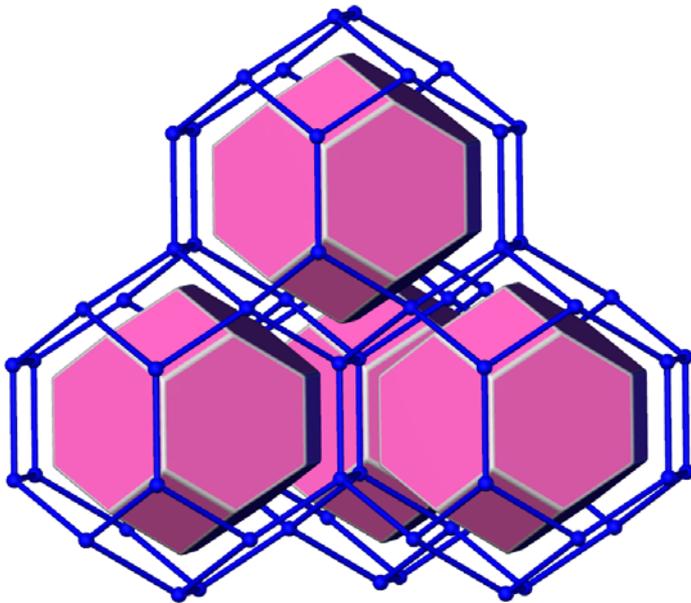
MD Foster and MMJ Treacy, A Database of Hypothetical Zeolite Structures: <http://www.hypotheticalzeolites.net>

For questions regarding the website please email: "martin" at "foster" dot "nu"

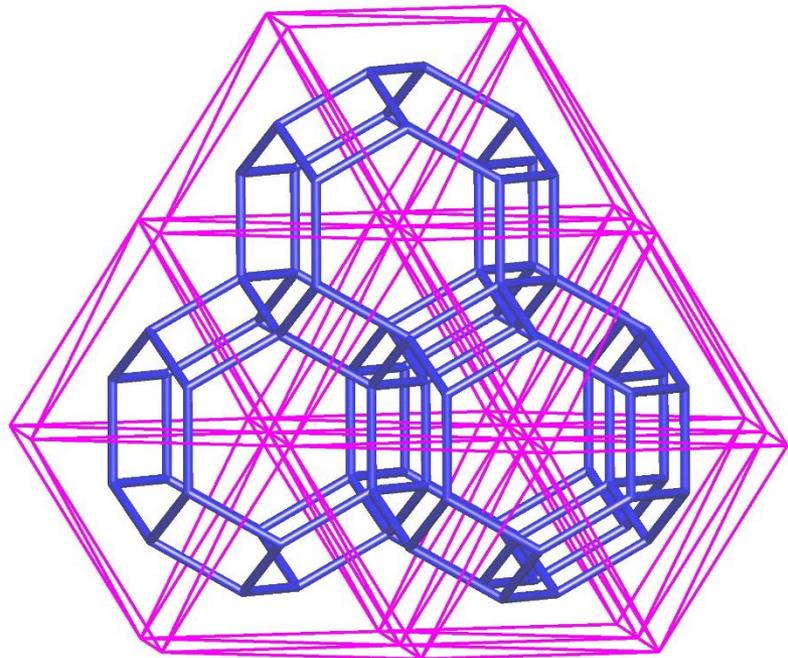
[Link to the old webpages](#)

Гипотетические:  
2 175 685 типов  
каркасов

# Тайлинг



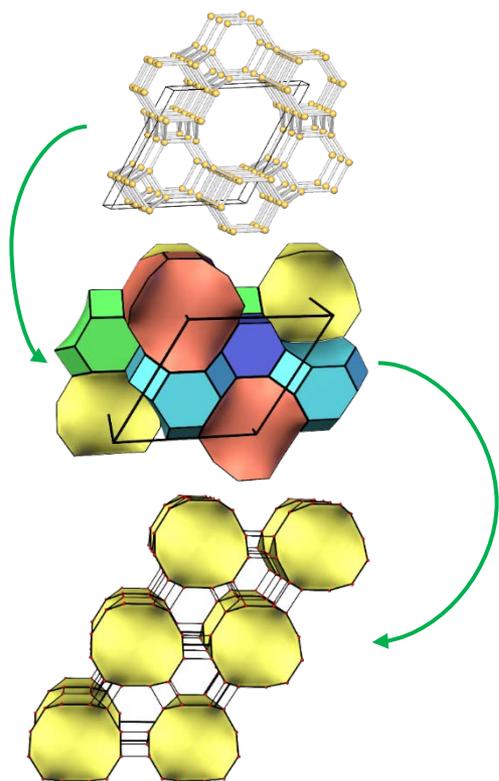
Топологический мотив каркаса  
содалита sod и его тайлинг  
 $24/14:[4^6.6^8]$



Дуальная сетка тайлинга  
 $24/14:[4^6.6^8] - \text{всу-х}$

# Теория самосборки кристаллических структур

## Новый метод анализа цеолитных каркасов



**232** цеолита

Строительные единицы	Цеолиты, построенные из одного типа тайлов
t-afo	BPH, AFS
t-ato	OFF
t-can	ERI, LIO, LTL, OFF
t-grc	KFI
t-hpr	AEI, AFT, AFX, CHA, EMT, FAU, GME, KFI, SAV, SFW, TSC
t-kaa	LTL
t-lau	LAU
t-lio	FAR
t-opr	MER, TSC
t-toc	FAU, LTA, EMT, IFY

Выработка критериев оценки «правильности» строения базовой сетки в структуре гипотетических цеолитов

Анализ геометрико топологических свойств базовой сетки реальных цеолитов

Построение базовой сетки для реальных цеолитов

- локальный тип связывания тайлов друг с другом:



- координационные числа каркасообразующих тайлов:



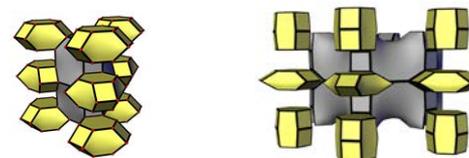
- координационные полиэдры:



- геометрическая форма тайлов:



- вторичные тайлы:



313 592  
базы данных  
Deem и BRONZE

1608 построены  
из одного типа  
тайлов

57 каркасов,  
соответствующих  
их критериям

ERI  $P6_3/mmc$

$[4^6.6^2]+[4^6.6^5]+[4^{12}.6^5.8^6]$

2 6 6 3

194\_5\_3713581  
 $P6_3/mmc$

$2[4^6.6^2]+5[4^6.6^5]+[4^{12}.6^5.8^6]+[4^{12}.6^{14}.8]$

5 15 15 6

Face symbol:

$[4^6.6^2]$



$[4^6.6^5]$



TILES

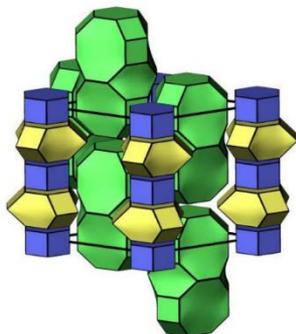
$[4^{12}.6^5.8^6]$



V, E, F: (12, 16, 8) (18, 27, 11) (42, 63, 23)  
Symmetry:  $\bar{3}m$   $6m2$   $6m2$   
Wyckoff:  $2a$   $2b$   $2d$   
Label: t-hpr t-can t-eri

Special features: *simple*

TILING



iza-structure.org

Face symbol:

$[4^6.6^2]$



$[4^6.6^5]$



TILES

$[4^{12}.6^5.8^6]$



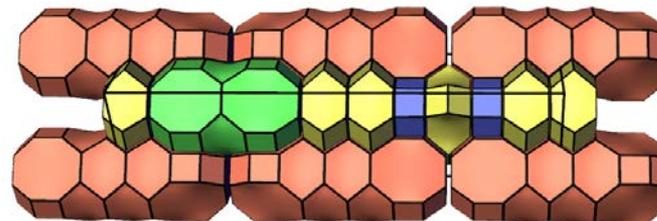
$[4^{12}.6^{14}.8]$



V, E, F: (12, 18, 8) (18, 27, 11) (42, 63, 23) (58, 86, 30)  
Symmetry:  $4f$   $2d, 4f$   $2d i$   $2a$   
Wyckoff:  $3m$   $\bar{6}m2, 3m$   $\bar{6}m2$   $\bar{3}m$   
Label: t-hpr t-can t-eri t-new

Special features: *simple*

TILING



# Поиск потенциальных структуронаправляющих агентов (SDA)

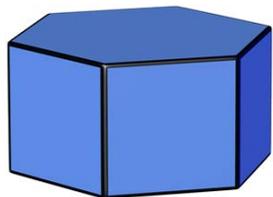
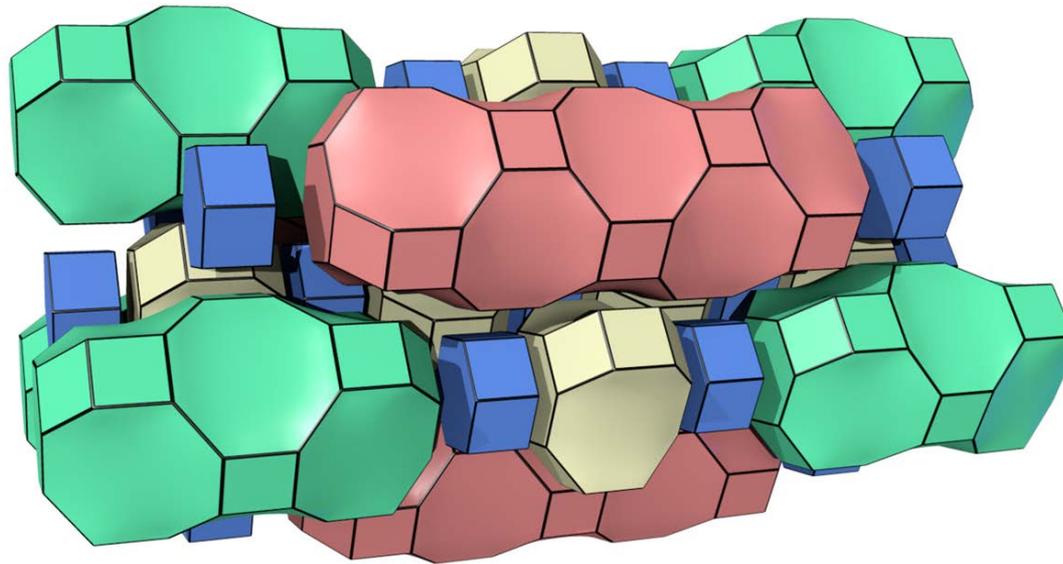
ГЕОМЕТРИЧЕСКИЕ ПАРАМЕТРЫ  
ПОЛОСТИ

```
graph TD; A[ГЕОМЕТРИЧЕСКИЕ ПАРАМЕТРЫ ПОЛОСТИ] --> B[ТТМ КОЛЛЕКЦИЯ]; B --> C[ПОТЕНЦИАЛЬНЫЕ SDA];
```

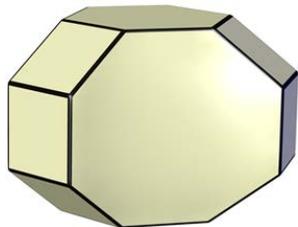
ТТМ КОЛЛЕКЦИЯ

ПОТЕНЦИАЛЬНЫЕ SDA

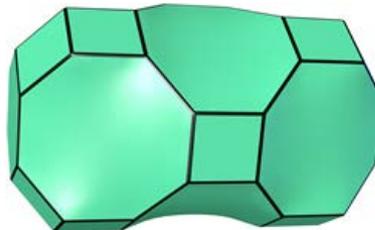
# Анализ полостей каркаса ISC-2



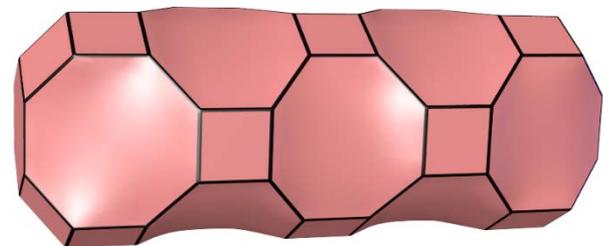
*t-hpr*



*t-gme*

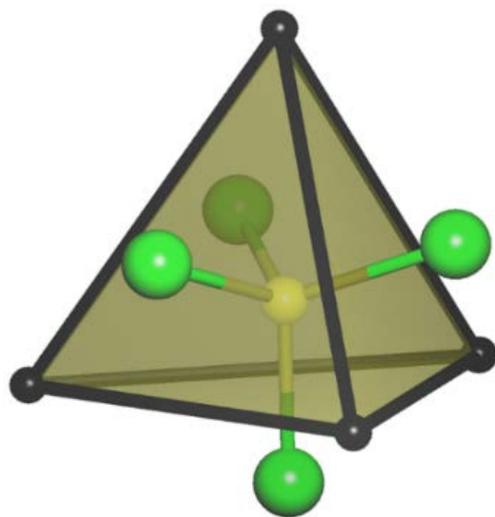


*t-aft*

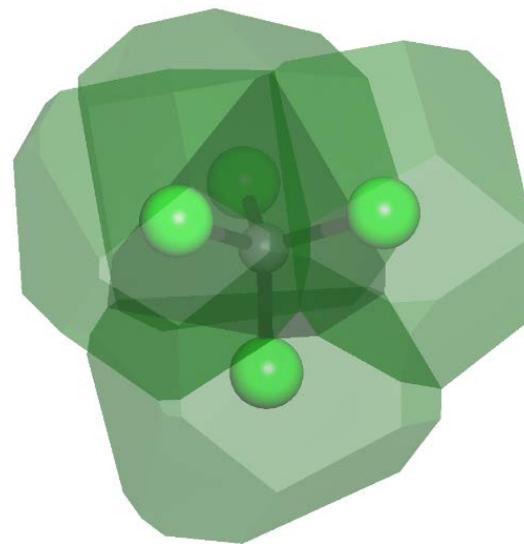


*t-isc-2*

# Разбиение пространства на полиэдры Вороного

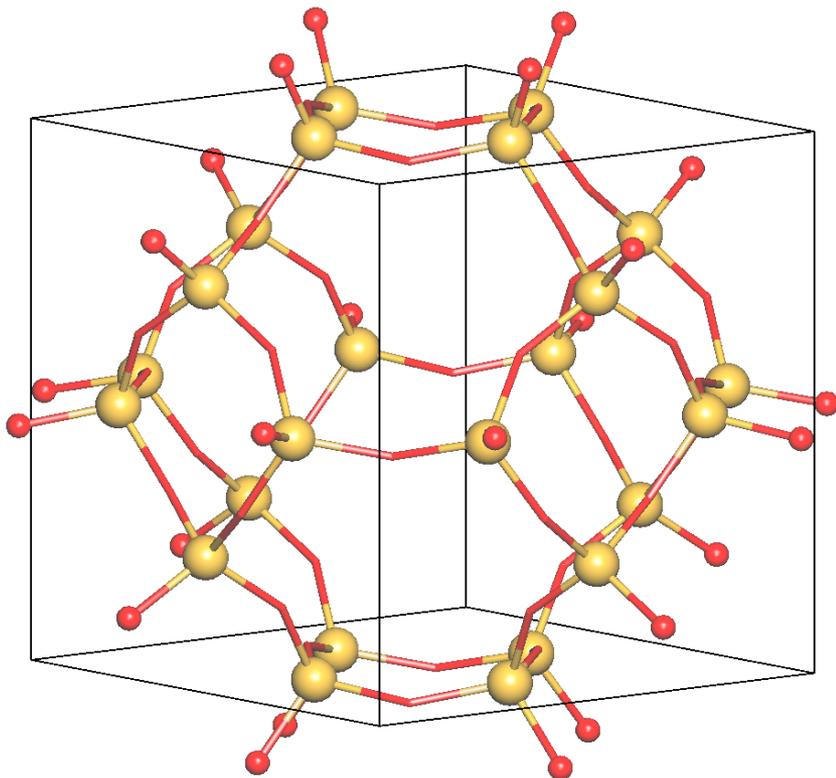


Полиэдр Вороного для атома в тетраэдрическом окружении  $\text{CCl}_4$



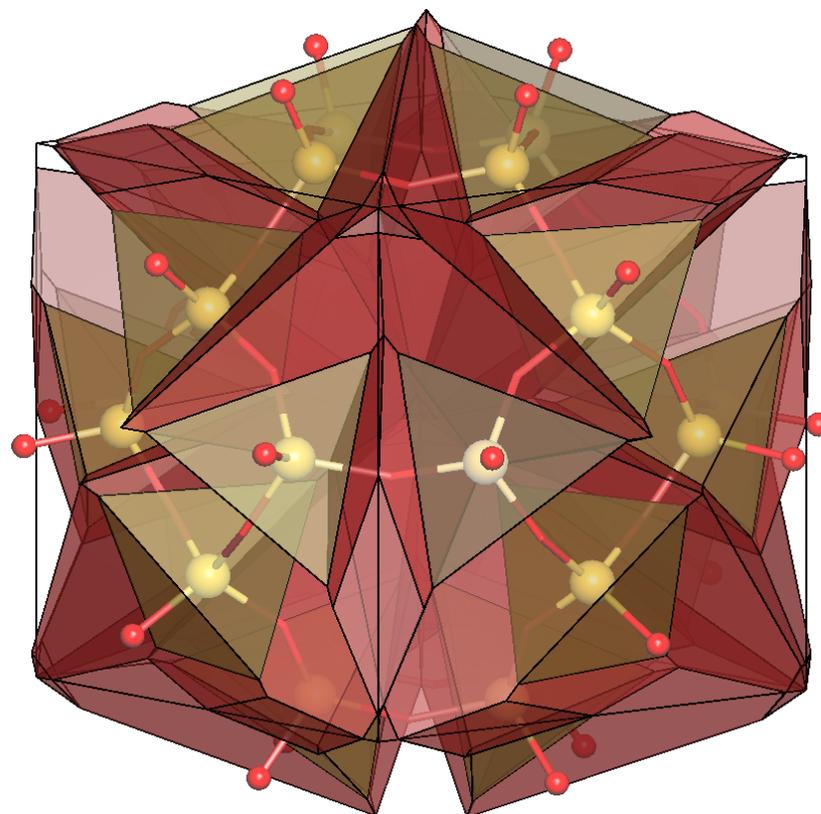
Молекулярный полиэдр Вороного в кристаллической структуре

# Сетка Вороного



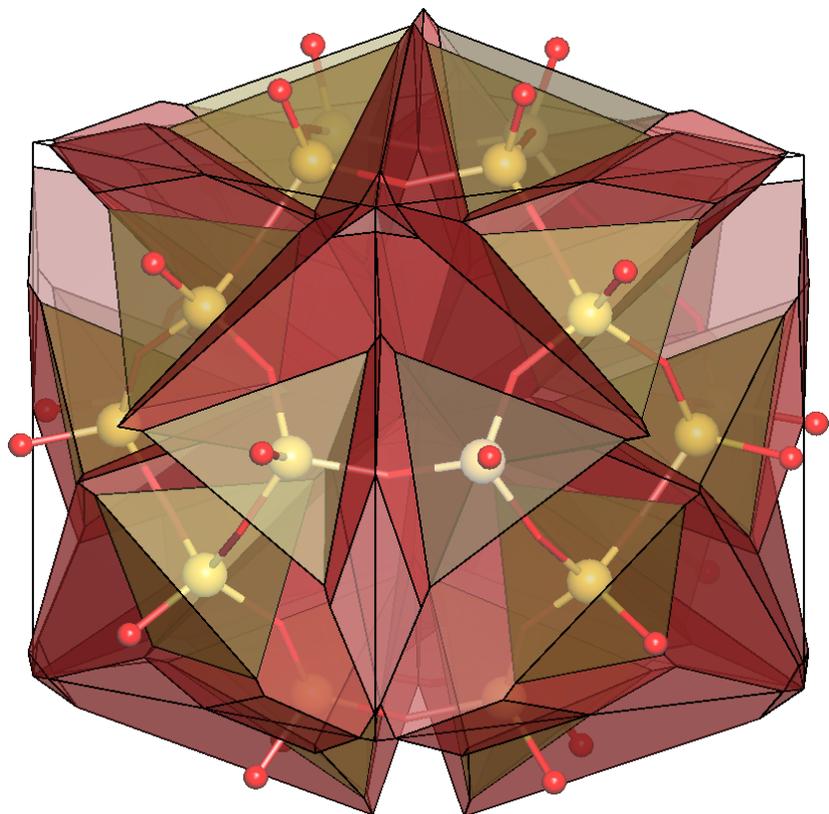
Алюмосиликатный каркас в структуре  
содалита  $\text{Na}_8(\text{Al}_6\text{Si}_6\text{O}_{24})\text{Cl}_2$

Loens, J. and Schulz, H. *Acta Crystallogr.*, 23, 434-436 (1967)

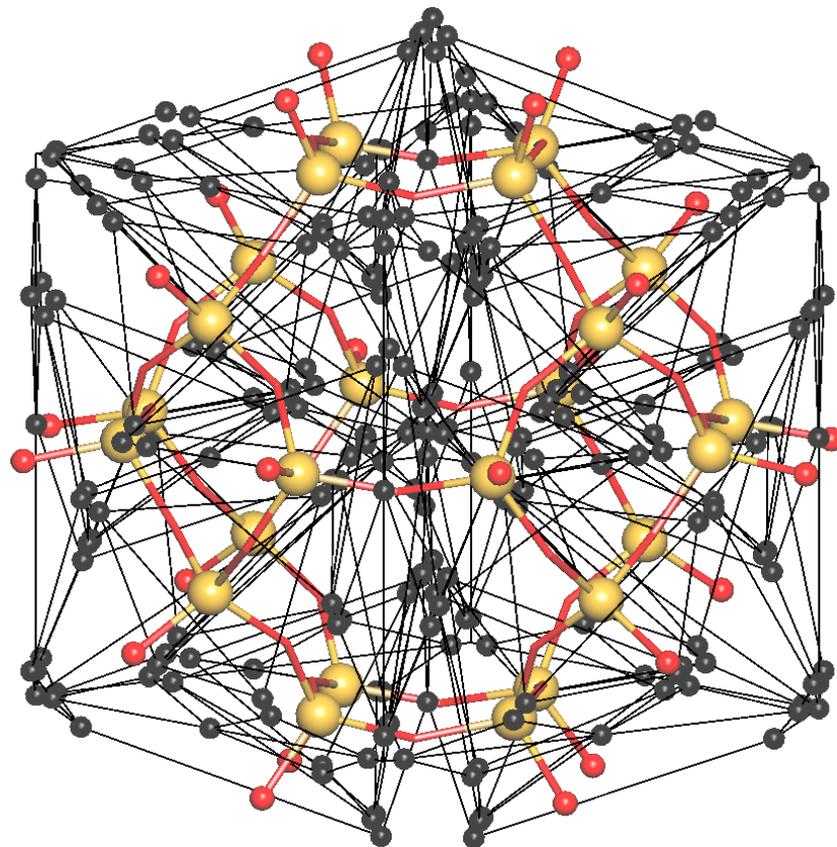


Полиэдры Вороного для атомов  
алюмосиликатного каркаса

# Сетка Вороного

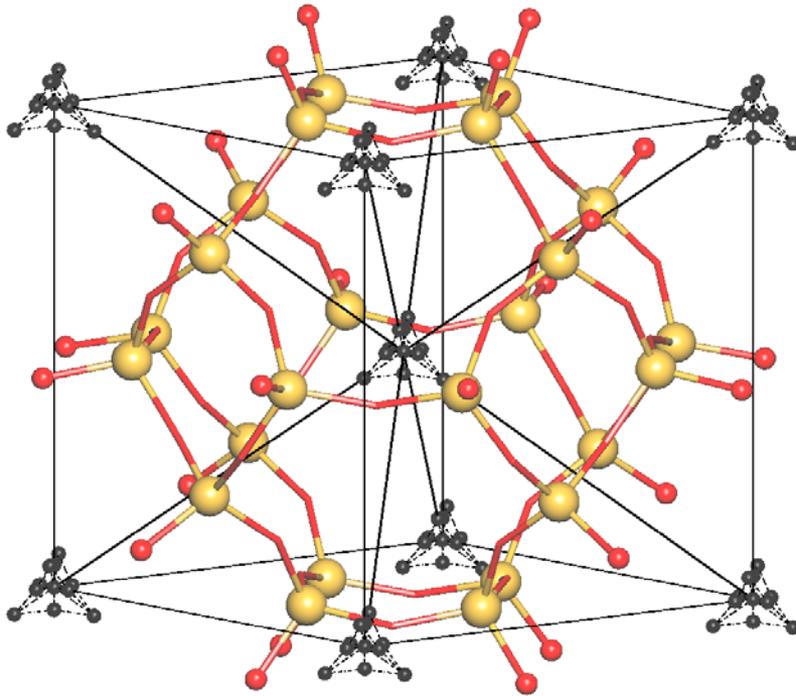


Разбиение Вороного

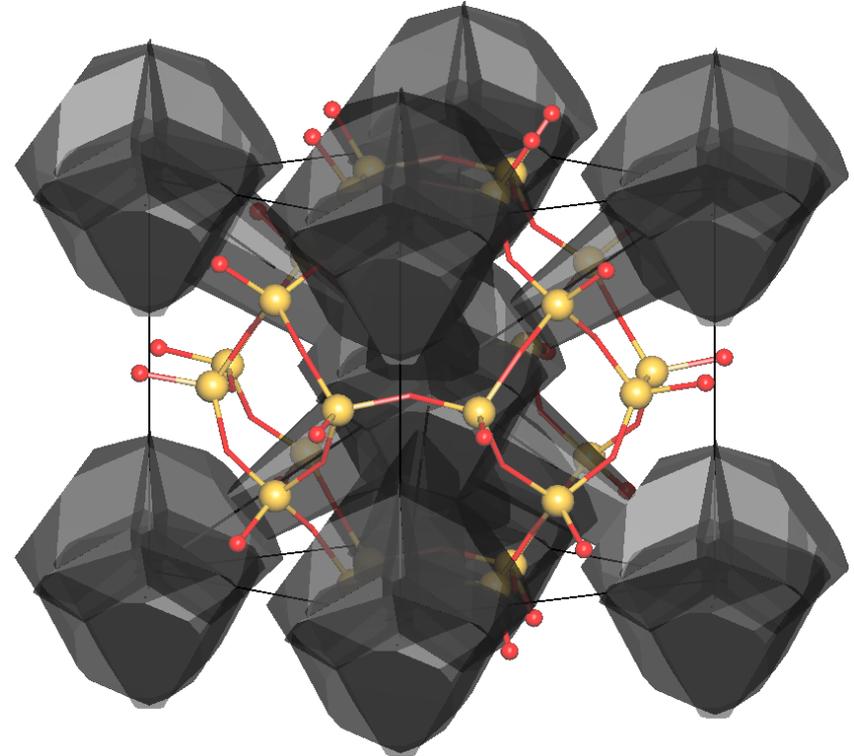


Сетка Вороного для  
содалитного каркаса

# Расчет геометрических параметров каналов



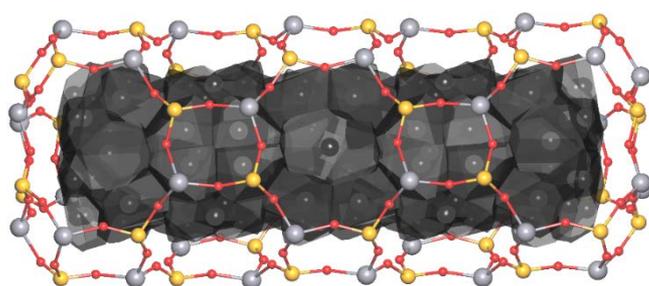
Подсетка 3-периодичной системы  
широчайших каналов



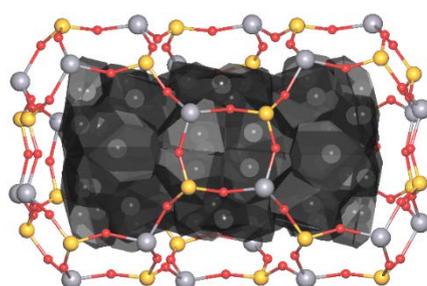
Полиэдры Вороного для узлов  
подсетки системы каналов

$V_{channel} = 210 \text{ \AA}^3$	$V_{channel}/V_{cell} = 0.32$
-----------------------------------	-------------------------------

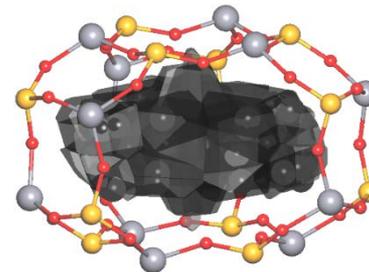
# Геометрические параметры полостей каркаса ISC-2



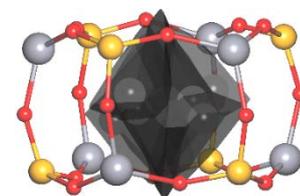
*isc-2*



*aft*



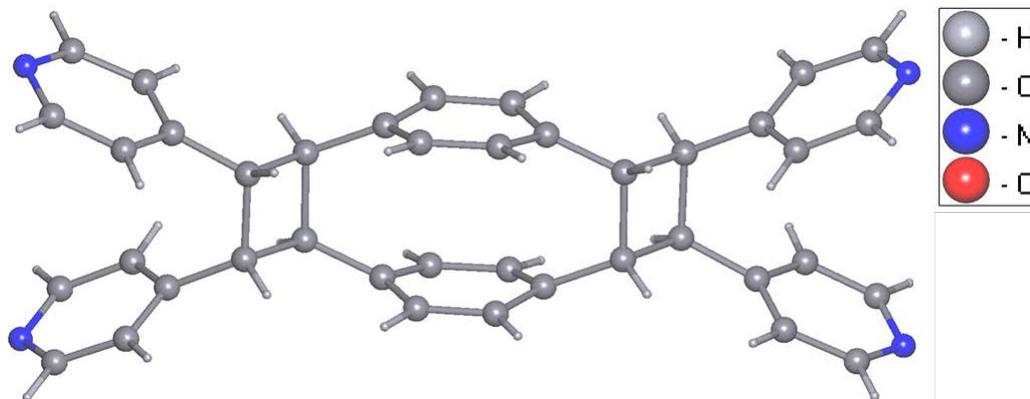
*gme*



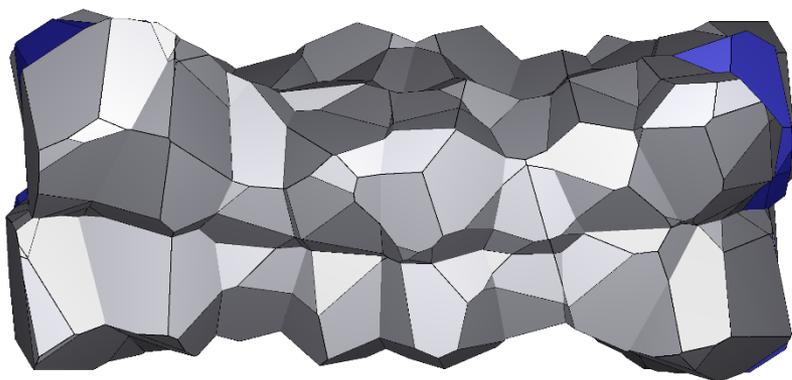
*hpr*

	<i>isc-2</i>	<i>aft</i>	<i>gme</i>	<i>hpr</i>
$V, \text{\AA}^3$	844	469	122	20
$V^*, \text{\AA}^3$	1006	596	125	22
$G_3$	0.174	0.104	0.096	0.080
$J_x/J_z$	0.169	0.408	0.655	0.781
$J_y/J_z$	0.992	0.982	0.733	0.795

# Потенциальный структуронаправляющий агент для синтеза цеолита ISC-2



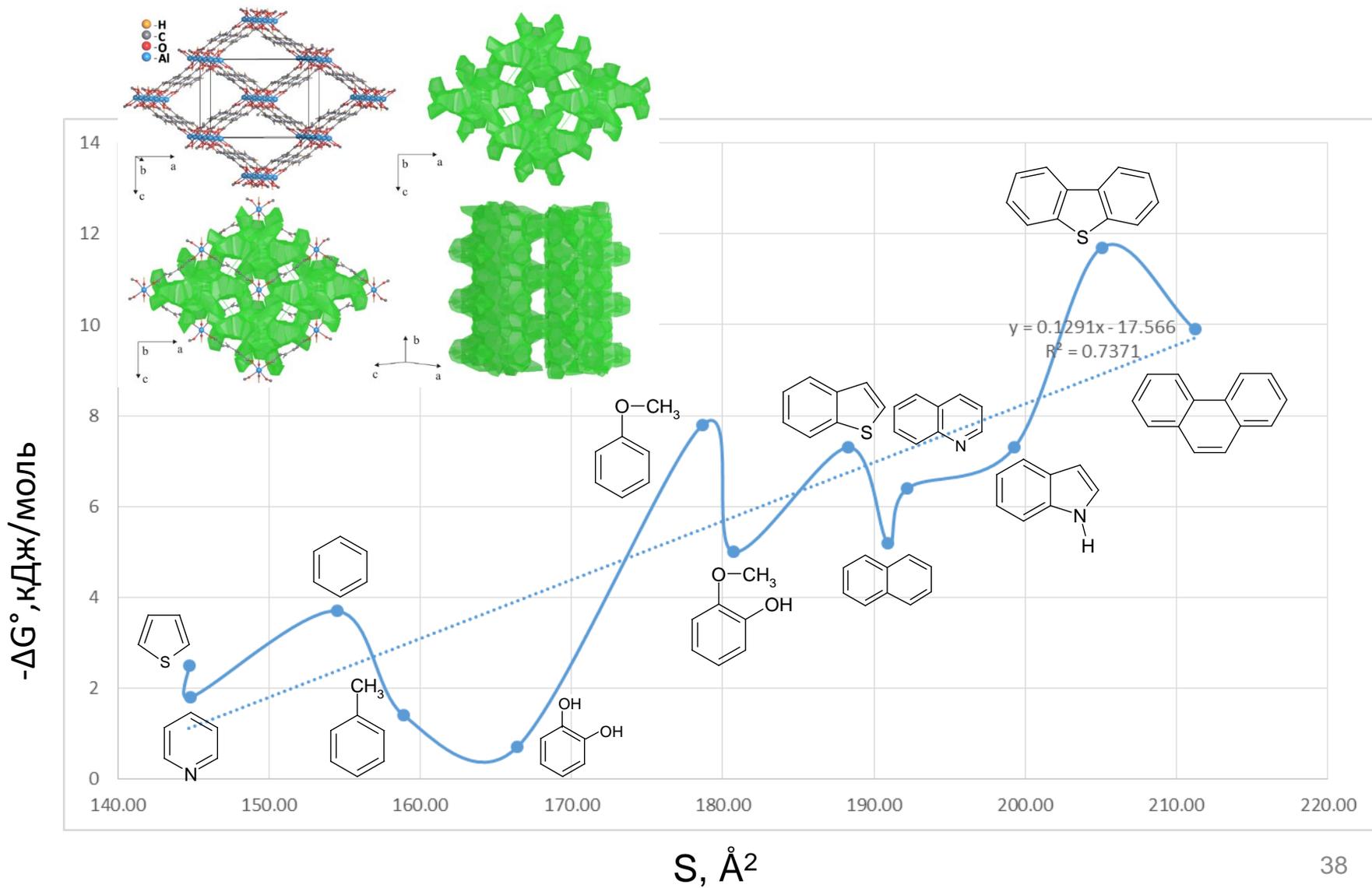
3,4,11,12-тетракис(4-пиридил)пентацикло  
(12.2.2.2<sup>6</sup>,9<sup>0</sup>.2,5<sup>0</sup>.10,13<sup>1</sup>)икоса-1(16),6,8,14,17,19-гексан



Молекулярный полиэдр Вороного-Дирихле

	<b>SDA</b>	<b>isc-2</b>
$V, \text{Å}^3$	712	844
$G_3$	0.152	0.174
$J_x/J_z$	0.215	0.169
$J_y/J_z$	0.966	0.992

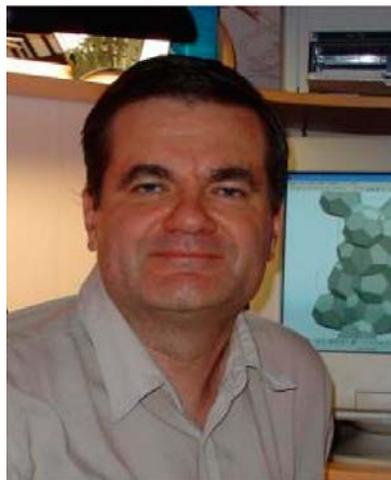
# Корреляции «геометрия пор-сорбция»







Давиде Прозерпио  
Миланский университет



Владислав Блатов



Александр Шевченко



Александр Вировец  
Институт неорганической химии (Новосибирск)



Евгения Пересыпкина



Андрей Голов



THE MINISTRY OF EDUCATION AND SCIENCE  
OF THE RUSSIAN FEDERATION

Megagrant 14.B25.31.0005

**Благодарю за внимание!**