XtalOpt User Guide Release 11.0

Generated by Doxygen 1.8.13

Contents

1	Xtal	Opt Tute	orial	1
	1.1	Launch	n XtalOpt	1
	1.2	Enter of	composition and restraints	2
		1.2.1	Chemical Composition	2
		1.2.2	Cell Parameters	2
		1.2.3	Interatomic Distances (IAD)	3
		1.2.4	Molecular-Unit Builder	3
		1.2.5	Random Spacegroup Generator	4
		1.2.6	Mitosis	4
	1.3	Optimi	zer setup	5
		1.3.1	VASP	5
		1.3.2	GULP	6
		1.3.3	PWscf	7
		1.3.4	CASTEP	8
		1.3.5	SIESTA	9
	1.4	Queue	setup	10
		1.4.1	Using a remote PBS cluster	11
		1.4.2	Using a remote SGE cluster	13
		1.4.3	Using a remote SLURM cluster	14
		1.4.4	Using a remote LSF cluster	15
		1.4.5	Using a remote LoadLeveler cluster	17
		1.4.6	Running optimizations locally	18
	1.5	What is	s written to the local directory?	18
	1.6	Search	Settings	19
	1.7	"Begin	"	20
	1.8	Monito	r progress	21
		1.8.1	View trends	22
		1.8.2	View Crystals in Avogadro2	23
		1.8.3	Plot a Theoretical XRD Pattern	24
	1.9	Comm	and Line Interface	25

1 XtalOpt Tutorial

2	Savi	ing and	Resuming Sessions in XtalOpt	26
	2.1	How to	save your session	27
	2.2	How to	resume your session	27
3	Opti	mizatio	n Schemes	27
	3.1	Overvie	ew: What are optimization schemes, and why use them?	27
		3.1.1	In a nutshell	27
		3.1.2	More details	28
	3.2	Optimiz	zation scheme user interface	29
		3.2.1	Optimization step list	29
		3.2.2	Add new optimization step	29
		3.2.3	Remove current optimization step	30
		3.2.4	Select template	30
		3.2.5	Template editor	30
		3.2.6	Save scheme	30
		3.2.7	Resume scheme	30
	3.3	How to	build an optimization scheme?	30
	3.4	How to	save an optimization scheme for later?	31
	3.5	How to	load an optimization scheme?	31
	3.6	What is	saved?	31
	3.7	Sugges	stions for optimization schemes	31
		3.7.1	Crystals (XtalOpt)	31

1 XtalOpt Tutorial

1.1 Launch XtalOpt

Simply run the "XtalOpt" executable (or in MacOS, open the XtalOpt.app file).

1.2	Enter	composition	and	restraints
-----	-------	-------------	-----	------------

		S	tructure Limit	S Optimization Settings	Search Settings Progress	Plot Log			
Composition					Mitosis				
Empirical Fo	rmula: 02 Ti	1			Use Mitosis	Export su	bcells?		
					# of Divisions:				
Formula	Units: 6				a: 🗘	b:	¢ c:	0	
Symbol	Z	#	Mass	Min. Radius	Space-group Generation				
0	8	2	15.9994	0.264					
Ti	22	1	47.867	0.64	Initialize with RandSpg?			Spg Options	
Internationalis Di					Linit Call Decemptors				
nteratomic Di	istances				Unit Cell Parameters				
Use Cust	tom Interatom	ic Distances				Minimum		Maximum	
Atom '	1 /	Atom 2		Min. IAD	Length A (Å):	1.00000	0	20.00000	0
					Length B (Å):	100000		20.00000	
					Longth D (A).	1.00000		20.00000	Ľ
🗸 Use Scale	ed Interatomic	Distances		Scale factor: 0.40 * radii	Length C (Å):	1.00000	0	20.00000	0
			м	linimum radius: 0.25 Å	Angle a (°):	60.00000	0	120.00000	0
Check IA	D Post-Optim	ization							
Check IA	D Post-Optim	ization			Angle β (°):	60.00000	٢	120.00000	0
Check IA	D Post-Optim lolecular Units	ization)Or #	Geometry Distance	Angle β (°): Angle γ (°):	60.00000 60.00000	0	120.00000	0 0
Check IA	D Post-Optim Iolecular Units	ization Neight	oor #	Geometry Distance	Angle β (°): Angle γ (°): Volume (ų/FU):	60.00000 60.00000 1.00		120.00000 120.00000 500.00	
Check IA Create M Center	D Post-Optim Iolecular Units	ization	oor #	Geometry Distance	Angle β (°): Angle γ (°): Volume (Å ³ /FU): Fixed volume	60.00000 60.00000 1.00	0 0 0	120.00000 120.00000 500.00	
Check IA	D Post-Optim	ization	oor #	Geometry Distance	Angle β (°): Angle γ (°): Volume (Å ³ /FU): Fixed volume	60.00000 60.00000 1.00		120.00000 120.00000 500.00	
Check IA	D Post-Optim olecular Units #	Neight	oor #	Geometry Distance	Angle β (°): Angle γ (°): Volume (Å ³ /FU): Fixed volume Volume (Å ³ /FU):	60.00000 60.00000 1.00		120.00000 120.00000 500.00	0
Check IA Create M Center	D Post-Optim Iolecular Units	Neight	oor #	Geometry Distance	Angle β (°): Angle γ (°): Volume (Å ³ /FU): Fixed volume Volume (Å ³ /FU):	60.00000 60.00000 1.00 500.00000		120.00000 120.00000 500.00	© © ©

The interface opens to the "Structure Limits" tab, shown above.

1.2.1 Chemical Composition

We will use a 6 formula unit supercell of titanium dioxide for this tutorial, so enter "Ti1 O2" for the "Empirical Formula". Under the "Formula Units" we can type 6. If the user wants to search multiple formula units, they would only have to type in the range, or individual numbers they wish to search (e.g., 1-6; or 1-2, 4, 6).

1.2.2 Cell Parameters

We will assume that we know nothing about the system and use very loose restraints (however, note that a search is much more effective if chemically reasonable restraints are used). Set all cell length minima to 1 angstrom and maxima to 20 angstroms. Constrain the angles to be between 60 and 120 degrees, and the volume from between 1 and 500 cubic angstroms. (Note that due to the angle adjustment described in CPC, 2011, 182, 372-387, 60-120 degrees is the largest range of cell angles that XtalOpt will generate.) Furthermore, the volume of the cell can be fixed, so that all cells generated will have the exact same volume.

1.2.3 Interatomic Distances (IAD)

There are now two different kinds of interatomic distances available: scaled interatomic distances and custom interatomic distances. If "Use Scaled Interatomic Distances" is checked, the covalent radii of the elements will be multiplied by the "Scale factor", and any radii below the "Minimum radius" will be set to the "Minimum Radius." The minimum interatomic distance, then, between pairs of atoms in this setup is the sum of their radii. For our example, check the "Use Scaled Interatomic Distances" checkbox and set "Scale factor" to 0.40 and "Minimum radius" to 0.25.

Custom interatomic distances is an alternative option. If the "Use Custom Interatomic Distances" box is checked instead, the user can specify the minimum interatomic distance between every pair of atom types in the table below the check box.

Finally, a checkbox labelled "Check IAD Post-Optimization" is also now available. If this box is checked, the interatomic distances are checked after the optimization is complete, and if any structures fail the interatomic distance check, they will be marked as failed structures.

1.2.4 Molecular-Unit Builder

If the user chooses to define specific molecular units, this option will allow them to do so. Once the chemical composition has been defined, the user can choose the center and neighbor atoms to be set as single-center molecules. For example, with the 6 formula unit TiO2, we could set 6 Ti atoms to be the center of 2 O atoms in a linear geometry. We can also decide what the interatomic distance between the center and neighboring atoms will be. If there are left over atoms, they will be placed randomly after the molecular units are added.

The molecular units are only used in the initial generation.

1.2.5 Random Spacegroup Generator

		Ti1	02		
	Space Group	Formula Units Possible	Allow randSpg?	Min xtals per	FU
1	P 1	16		0	٢
2	P -1	16		0	٢
3	P 1 2 1	16		0	٢
4	P 1 21 1	16		0	٢
5	C121	16		0	٢
6	P 1 m 1	16		0	٢
7	P1c1	16		0	٢
8	C 1 m 1	16		0	٢
9	C1c1	16		0	٢
10	P 1 2/m 1	16		0	٢
11	P 1 21/m 1	16		0	٢
12	C 1 2/m 1	16		0	٢
13	P 1 2/c 1	16		0	٢
14	P 1 21/c 1	16		0	٢
15	C 1 2/c 1	16		0	٢
16	P 2 2 2	16		0	٢
	Select all	Decelect ell	Incoment All	Deeromen	

With the implementation of RandSpg, the initial generation of structures can be created by using specific spacegroups (or a variety of spacegroups). The choice for spacegroups will be limited by the chemical composition and number of formula units.

Spacegroup constraints can only be used in the initial generation. Molunit and randSpg cannot be used simultaneously.

For more information on randSpg, see this paper.

1.2.6 Mitosis

For structures with large unit cells/large number of atoms, one can use "Mitosis" to increase the local order of the initially generated structures. This will create a small subcell, then multiply the subcell in each direction to fulfill the stoichiometry and unit cell size. "# of Divisions" defines how many subcells will be replicated, based upon the user defined number of "Formula Units", and the a, b, c determine in which direction the cells are replicated.

1.3 Optimizer setup

XtalOpt currently supports the VASP, GULP, PWscf, CASTEP, and SIESTA codes for performing geometry optimizations. Each is detailed in its own section below.

Be aware that program installation is different, and it is almost certain that the submit files included with these schemes will not work on any cluster other than the Zurek group's "parity" cluster at SUNY Buffalo's Center for Computational Research. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance for information about MPI, executable locations, etc. Perhaps the easiest method to find the correct submit script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

1.3.1 VASP

•	<i>⊷</i> ∧taiopt
	Structure Limits Optimization Settings Search Settings Progress Plot Log
Queue PBS Oconfigure	\$filename\$
Optimizer VASP Configure	# output options LMARE = .FALSE. # write or don't write WAVECAR LCRARG = .FALSE. # write or don't write CEG and CHOCAR LELF = .FALSE. # write ELF
Template INCAR Cptimization 2 Optimization 2 Optimization 3 Optimization 5 Optimization 5 Optimization 6	<pre>ionic selaxation WH NO 2 # 2-econjucate gradient, 1=Newton like WH NO 2 # 2-econjucate gradient, 1=Newton like USIF 2 # 3-reclax verything, 2-relax ions only, 4=keep volume fixed # precision parameters DDIFTG = Is-2 # usually 10 * DDIFF DDIFTG = Is-2 # usually 10 * DDIFF PRCE = reclassation USERER = 5 # 5 = setredon, 1N = Nethfessel HNUTT = %user1% # cutoff energy PSTST = %user2%</pre>
Add Remove Help Save Opt Scheme Load Opt Scheme	
user1:	
user2: 0	
user4:	
Save Session Resume stored session	Optimized: 0 Running: 0 Failures: 0 Begin

On the next tab, load the optimization scheme by clicking the "Load Opt Scheme" button and selecting the "samples/vasp-xtalopt.scheme" file that is distributed with the source code. If you do not have a copy of the source code, the scheme file can be obtained by clicking here.

For more details on optimization schemes, see Optimization Schemes.

After loading the optimization scheme, XtalOpt will prompt for the POTCAR files to use. Select files appropriate for the prompted atom. XtalOpt will construct the POTCAR files on the local computer, and then copy them over to the cluster when the calculation is submitted. It is necessary to have the VASP POTCAR files for each atomic species located somewhere on the local computer. See the VASP manual for information on obtaining the POTCAR files.

Take a moment to look through each file for each optimization step. Notice that the INCAR template includes two user-specified values, %user2% and %user3% for the external pressure and the energy cutoff, respectively. By entering appropriate values in the "user2:" and "user3:" fields on the left, it is easy to update these values for all optimization steps.

Notice the other %keyword% values in the job.pbs templates. These are used to enter information that is specific to a search or structure when the actual input files are written prior to job submission. Click the "Help" button for a full listing of the available keywords.

XtalOpt expects VASP to use the default filenames, mainly POSCAR, CONTCAR, and OUTCAR.

Skip to next section.

1.3.2 GULP

				<i>io</i> 1	XtalOpt			
			Structure Limits	ptimization Settings	Search Settings	Progress	Plot Log	
Queue Optimizer Template Optimizat	Local C QULP C xtal.gin	Configure	opti conj conp ovich_minimiser b colic frac teocordsTraces species D -1.098 buck Ti Ti 3120.10.15 0 0 11087.30.23 Ti 0 10.15 0 0 1 0 15 Ti 0 1 0 15	fgs gnorm 0.5 haDeg% %betaDeg% % 40 5.25 15 40 30.22 15 40 12.59 15	gammaDegi			
		Hala						
Add. Save (user1:	. Remove	id Opt Scheme						

On the next tab we choose GULP for the local optimizer and enter a template for GULP to use. Select "GULP" as the "Optimizer" and "xtal.gin" as "Template". Next, fill out the text field on the right with the following template:

opti conj conj	p		
switch_minimi:	ser bfgs g	ynorm 0.5	
cell			
응a응 응b응 응C응	%alphaDec	g% %betaDeg%	%gammaDeg%
frac			
%coordsFrac%			
species			
Ti 2.196			
0 -1.098			
buck			
Ti Ti 31120.1	0.1540 5.	.25 15	
0 0 11782.7	0.2340 30	0.22 15	
Ti O 16957.5	0.1940 12	2.59 15	
lennard 12 6			
Ti Ti 1 0 15			
0 0 1 0 15			
Ti O 1 O 15			

Alternatively, one can load the scheme file distributed with the source code under samples/gulp-TiO-xtalopt.scheme. If the source code is not available, the scheme file can be obtained by clicking here.

For more details on optimization schemes, see Optimization Schemes.

Note the "%" surrounding various keywords. These will be replaced by the structure-specific data when the optimizer is invoked for each structure. Click "Help" to view all of the keywords available. The number of optimization steps can be modified with the "Add/Resume" buttons. The "user" fields in the lower left corner allow users to specify their own keyword/value pairs, which is useful for making changes to multiple optimization steps at once. We will only be using one optimization step in this tutorial.

XtalOpt expects GULP to use the following filenames:

gulp < xtal.gin > xtal.got

Skip to next section.

1.3.3 PWscf



On the next tab, load the optimization scheme that is distributed with the source code under the samples/ directory. The scheme that we want is named "pwscf-xtalopt.scheme". If the source code is not available, the scheme file can be obtained by clicking the "Original Format" link at the bottom of the page here.

For more details on optimization schemes, see Optimization Schemes.

Each PWscf input file will need to be edited to specify:

- 1. The pseudo_dir containing the pseudopotential files on the remote cluster, and
- 2. The pseudopotentials for each atom (under ATOMIC_SPECIES)

Take a moment to look through each file for each optimization step.

Notice the %keyword% values in the job.pbs templates. These are used to enter information that is specific to a search or structure when the actual input files are written prior to job submission. Click the "Help" button for a full listing of the available keywords.

XtalOpt expects PWscf to use the following filenames:

pw.x < xtal.in > xtal.out

Skip to next section.

1.3.4 CASTEP

	Structure Li	imits Optimization Settings Search Settin	gs Progress Plot Log	
Queue PBS	Configure Repercent 8B	LOCK LATTICE CART		
	Scellvector Scellvector	rlAngstrom%		
Optimizer CASTEP	Configure %cellVector %percent%E	r3Angstrom% NDBLOCK LATTICE_CART		
Template xtal.cell	\$percent %BJ	LOCK POSITIONS_FRAC		
Optimization 1	scoordsride spercent%E	C% NDBLOCK POSITIONS_FRAC		
Add Remov	e Help			
Save Ont Scheme	Load Ont Scheme			
Cuve opt Scheme	Loud Opt Outerine			
user1:				
10072				
user3:				
user4:				
Deven Operation	alayed energies	Outlation to Dur		Basia
Hesum	7 SUIEU SESSION	Optimized: U Huni	ing: 0 Failures: 0	Begin

On the next tab, load the optimization scheme that is distributed with the source code under the samples/ directory. The scheme that we want is named "castep-xtalopt.scheme". If the source code is not available, the scheme file can be obtained by clicking the "Original Format" link at the bottom of the page here.

For more details on optimization schemes, see Optimization Schemes.

It is important to note that CASTEP input files require the "%" character to define blocks. The percent character is special in the XtalOpt input template parser to define keywords (see below). To insert a literal "%" into the input, use percent%.

E.g. Specification of the fractional coordinate block in the .cell template should look like:

%block% POSITIONS_FRAC
%coordsFrac%
%endblock% POSITIONS_FRAC

Take a moment to look through each file for each optimization step.

Notice the %keyword% values in the job.pbs templates. These are used to enter information that is specific to a search or structure when the actual input files are written prior to job submission. Click the "Help" button for a full listing of the available keywords.

XtalOpt expects CASTEP to use the following filenames:

```
# XtalOpt will write xtal.cell, xtal.param
castep xtal
# CASTEP will create xtal.castep
```

Skip to next section.

1.3.5 SIESTA

				Structure Limits	Optimization Settin	gs Search Settings	Progress	Plot L	og
Queue	Local	Cor	nfigure	SystemName SystemLabel	%description xtal	8-%gen%x%id%			
Optimizer	SIESTA	Cor	nfigure	NumberOfAtoms NumberOfSpecies	%numAtoms% %numSpecies%				
Template	xtal.fdf		٥	%block% Chemicals	peciesLabel	•			
Optimizatio	on 1			<pre>%endblock% Chemic %block% PAO.Basis Ti 5 1.9</pre>	alSpeciesLabel	♥ Define Basis se	2		
				n=3 0 1 5.699466626162 1.0000000000000	E 93.95	5.20			
				n=3 1 1 5.699413394659 1.00000000000	E 95.47 94	5.20			
				n=4 0 2 6.099963989753 1.000000000000000 n=3 2 2	E 96.47 07 5.099 00 1.000 E 46.05	5.60 44363262274 0000000000 4.95			
				5.94327035784 1.00000000000 n=4 1 1 3.053659799385 1.00000000000 %endblock% PAO.B:	517 4.700 100 1.000 E 0.50 136 1000 1515	09988294302 00000000000 1.77			
				<pre>%block% PAO.Basis 0 Di %endblock% PAO.Basis</pre>	Sizes P sisSizes				
Add	Remov	/e	Help	XC.functional GGA XC.authors PBB					
Save C	Opt Scheme	Load Opt S	Scheme	LatticeConstant %block% LatticeVe %cells %endblock% Lattice	1 Ang ctors datrixAngstrom% ceVectors				
user1: user2:				<pre>%block% kgrid_Mom 3 0 0 0.5 0 3 0 0.5 0 0 3 0.5 %endblock% Kgrid_</pre>	khorst_Pack Monkhorst_Pack				
user3:				MeshCutoff	100.0 Ry				
user4:				MaxSCFIterations DM.MixingWeight	1000 0.05				
Save Sessio	on Resum	e stored sessi	ion		Optimi	zed: 0 Running:	0 Fai	ilures: 0	Begin H

On the next tab we choose SIESTA for the local optimizer and enter a template for SIESTA to use. Select "SIESTA" as the "Optimizer" and "xtal.fdf" as "Template".

Next, fill out the text field on the right with the following template:

SystemName SystemLabel	%description%- xtal	%gen%x%id%
NumberOfAtoms NumberOfSpecies	%numAtoms% %numSpecies%	
%block% ChemicalSpec %chemicalSpeciesLabe %endblock% ChemicalS	iesLabel 1% SpeciesLabel	
%block% PAO.Basis		# Define Basis set
n=3 0 1 E 5.69946662616249 1.000000000000000	93.95	5.20
n=3 1 1 E 5.69941339465994 1.000000000000000	95.47	5.20
n=4 0 2 E 6.09996398975307 1.000000000000000	96.47 5.09944 1.00000	5.60 1363262274 000000000
n=3 2 2 E 5.94327035784617 1.0000000000000000	46.05 4.70009 1.00000	4.95 9988294302 000000000
n=4 1 1 E 3.05365979938936 1.000000000000000	0.50	1.77
<pre>%endblock% PAO.Basis %block% PAO.Basis</pre>	3	
%endblock% PAO.Basis	Sizes	
XC.functional GGA XC.authors PBE		

LatticeConstant 1 Ang %block% LatticeVectors %cellMatrixAngstrom% %endblock% LatticeVectors %block% kgrid_Monkhorst_Pack 3 0 0 0.5 0 3 0 0.5 0 0 3 0.5 %endblock% Kgrid_Monkhorst_Pack MeshCutoff 100.0 Rv MaxSCFIterations 1000 DM.MixingWeight 0.05 DM.NumberPulay DM.Tolerance 1.d-4 SolutionMethod diagon SpinPolarized true LongOutput true MD.TypeOfRun ca 1000 MD.NumCGsteps MD.VariableCell true 0.01 eV/Ang #0.005 eV/Ang MD.MaxForceTol WriteForces true WriteCoorCerius true WriteCoorXmol false WriteDenchar true WriteMullikenPop 1 UseSaveData true #Diag.ParallelOverK true AtomicCoordinatesFormat Fractional

%block% AtomicCoordinatesAndAtomicSpecies %atomicCoordsAndAtomicSpecies% %endblock% AtomicCoordinatesAndAtomicSpecies

Or load the optimization scheme by clicking the "Load Opt Scheme" button and selecting the "samples/siesta-TiOxtalopt.scheme" file that is distributed with the source code. If the source code is not available, the scheme file can be obtained by clicking here.

For more details on optimization schemes, see Optimization Schemes.

After loading the optimization scheme, XtalOpt will prompt for the xtal.psf files to use. Select files appropriate for the prompted atom. XtalOpt will copy the individual files over to each structure directory when the calculation is submitted. See the SIESTA manual for information on obtaining the .psf files.

Notice the other %keyword% values in the xtal.fdf templates. These are used to enter information that is specific to a search or structure when the actual input files are written prior to job submission. Click the "Help" button for a full listing of the available keywords.

Note: that in the current implementation XtalOpt uses the Total Final Energy printed in the output to determine the fitness of a structure. If the user would like to use a different thermodynamic quantity for the fitness, please contact the XtalOpt developers.

XtalOpt expects SIESTA to use the following filenames:

siesta < xtal.fdf > xtal.out

1.4 Queue setup

XtalOpt currently supports using the PBS, SGE, SLURM, LSF, and LoadLeveler queuing systems on remote SS \leftrightarrow H-accessible clusters, as well as an internal local queue that manages calculations on the user's workstation. Each queueing interface is detailed in its own section below.

1.4.1 Using a remote PBS cluster

Host:	somehost.com	SSH Port 22
User:	xtalopt-user	
Working directory (Server):	/projects/thisrun	
Working directory (Local):	/calcs/thatrun	
Description:	Imajob	
Path to qsub:	qsub	
Path to qdel:	qdel	
Path to qstat:	qstat	
Queue refresh interval:	10 sec	
Clean remote directories	when finished	_
		Cancel OK

Select "PBS" from the list of Queues, and then click the "Configure..." button. A new window will prompt for:

- · host: The hostname of the PBS cluster's head node
- · user: The username used to log into the cluster
- Working directory (Server): A directory that is readable/writable by "user" on the cluster, used when performing optimizations.
- Working directory (Local): A directory that is readable/writable by the current user on the local computer. This is where the final structures and resume files are written.
- · Description: Used for the %description% keyword in input templates.
- Path to qsub: Where to find the qsub executable on the remote cluster. Note that if qsub is in the cluster's \$PATH, setting this to just 'qsub' will work.
- Path to qdel: Where to find the qdel executable on the remote cluster. Note that if qdel is in the cluster's \$PATH, setting this to just 'qdel' will work.
- Path to qstat: Where to find the qstat executable on the remote cluster. Note that if qstat is in the cluster's \$PATH, setting this to just 'qstat' will work.
- "Clean remote directories when finished": will remove all of the generated files from the cluster. Only the files on the local computer will be kept. If you do not want this to occur, make sure to uncheck this option.

A new template, "job.pbs" is added to the list of available templates. This is the job submission script for PBS. This script should roughly follow this design:

```
#/bin/bash
#PBS -1 nodes=1:ppn=8
#PBS -o ../%gen%x%id%-%optstep%.out
#PBS -e ../%gen%x%id%-%optstep%.err
#PBS -N %description%-%gen%x%id%-%optstep%
###Include this for XtalOpt scripts!###
export PBS_O_WORKDIR=%rempath%
# Change to structure's working directory, copy input files to node's scratch dirs:
for node in `cat $PBS_NODEFILE | sort | uniq'; do
rsh $node "cp $PBS_O_WORKDIR/* $PBSTMPDIR/;";
done
# Move to the scratch directory
cd $PBSTMPDIR
echo "running in directory $PBSTMPDIR"
# Set any environment variables needed for the optimizer/MPI here:
# Run optimizer, be sure to use the filenames that XtalOpt expects.
# See the template menu in XtalOpt and the example templates in the
# samples/ directory of the XtalOpt sources.
# Don't forget to clean up after MPI if needed!
// Print files from each node
for node in `cat $PBS_NODEFILE | sort | uniq`; do
echo "$node:"
rsh $node "ls -l $PBSTMPDIR"
done
# Copy back results from master node's scratch directory
cp $PBSTMPDIR/* $PBS_O_WORKDIR/
```

For more details on optimization schemes, see Optimization Schemes.

A handy trick for monitoring jobs outside of XtalOpt is to include the following line in job.pbs:

#PBS -N %description%-%gen%x%id%-%optstep%

This will name each job, for example, xtalSearch-3x4-2, where xtalSearch is a user-specified description of the search, and 3x4-2 means that it is the fourth structure in the third generation running its second optimization step.

Skip to next section.

1.4.2 Using a remote SGE cluster

lost:	somehost.com	SSH Port 22	0
Jser:	xtalopt-user		
Working directory (Server):	/projects/thisrun		
Working directory (Local):	/calcs/thatrun		
Description:	Imajob		
Path to qsub:	qsub		
Path to qdel:	qdel		
Path to qstat:	qstat		
Queue refresh interval:	1 sec		0
Clean remote directories	when finished		
		Cancel	

Select "SGE" from the list of Queues, and then click the "Configure..." button. A new window will prompt for:

- · host: The hostname of the SGE cluster's head node
- · user: The username used to log into the cluster
- Working directory (Server): A directory that is readable/writable by "user" on the cluster, used when performing optimizations.
- Working directory (Local): A directory that is readable/writable by the current user on the local computer. This is where the final structures and resume files are written.
- · Description: Used for the %description% keyword in input templates.
- Path to qsub: Where to find the qsub executable on the remote cluster. Note that if qsub is in the cluster's \$PATH, setting this to just 'qsub' will work.
- Path to qdel: Where to find the qdel executable on the remote cluster. Note that if qdel is in the cluster's \$PATH, setting this to just 'qdel' will work.
- Path to qstat: Where to find the qstat executable on the remote cluster. Note that if qstat is in the cluster's \$PATH, setting this to just 'qstat' will work.
- "Clean remote directories when finished": will remove all of the generated files from the cluster. Only the files on the local computer will be kept. If you do not want this to occur, make sure to uncheck this option.

A new template, "job.sge" is added to the list of available templates. This is the job submission script for SGE. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance or information about MPI, executable locations, etc. Perhaps the easiest method to find the correct SGE script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

For more details on optimization schemes, see Optimization Schemes.

Skip to next section.

1.4.3 Using a remote SLURM cluster

Host:	somehost.com	SSH Port 22	0
User:	xtalopt-user		
Working directory (Server):	/projects/thisrun		
Working directory (Local):	/calcs/thatrun		
Description:	Imajob		
Path to sbatch:	sbatch		
Path to scancel:	scancel		
Path to squeue:	squeue		
Queue refresh interval:	1		0
Clean remote directories	when finished		
		Cancel	ОК

Select "SLURM" from the list of Queues, and then click the "Configure..." button. A new window will prompt for:

- · host: The hostname of the SLURM cluster's head node
- · user: The username used to log into the cluster
- Working directory (Server): A directory that is readable/writable by "user" on the cluster, used when performing optimizations.
- Working directory (Local): A directory that is readable/writable by the current user on the local computer. This is where the final structures and resume files are written.
- Description: Used for the %description% keyword in input templates.

- Path to sbatch: Where to find the sbatch executable on the remote cluster. Note that if sbatch is in the cluster's \$PATH, setting this to just 'sbatch' will work.
- Path to scancel: Where to find the scancel executable on the remote cluster. Note that if scancel is in the cluster's \$PATH, setting this to just 'scancel' will work.
- Path to squeue: Where to find the squeue executable on the remote cluster. Note that if squeue is in the cluster's \$PATH, setting this to just 'squeue' will work.
- "Clean remote directories when finished": will remove all of the generated files from the cluster. Only the files on the local computer will be kept. If you do not want this to occur, make sure to uncheck this option.

A new template, "job.slurm" is added to the list of available templates. This is the job submission script for S \leftarrow LURM. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance or information about MPI, executable locations, etc. Perhaps the easiest method to find the correct SLURM script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

For more details on optimization schemes, see Optimization Schemes.

Skip to next section.

1.4.4 Using a remote LSF cluster

Host:	somehost.com	SSH Port 22
User:	xtalopt-user	
Working directory (Server):	/projects/thisrun	
Working directory (Local):	/calcs/thatrun	
Description:	Imajob	
Path to bsub:	bsub	
Path to bkill:	bkill	
Path to bjobs:	bjobs	
Clean remote directories	when finished	
		Cancel OK

Select "LSF" from the list of Queues, and then click the "Configure..." button. A new window will prompt for:

· host: The hostname of the LSF cluster's head node

- user: The username used to log into the cluster
- Working directory (Server): A directory that is readable/writable by "user" on the cluster, used when performing optimizations.
- Working directory (Local): A directory that is readable/writable by the current user on the local computer. This is where the final structures and resume files are written.
 - Description: Used for the description% keyword in input templates.
 - Path to bsub: Where to find the bsub executable on the remote cluster. Note that if bsub is in the cluster's \$PATH, setting this to just 'bsub' will work.
 - Path to bkill: Where to find the bkill executable on the remote cluster. Note that if bkill is in the cluster's \$PATH, setting this to just 'bkill' will work.
 - Path to bjobs: Where to find the bjobs executable on the remote cluster. Note that if bjobs is in the cluster's \$PATH, setting this to just 'bjobs' will work.
- "Clean remote directories when finished": will remove all of the generated files from the cluster. Only the files on the local computer will be kept. If you do not want this to occur, make sure to uncheck this option.

A new template, "job.lsf" is added to the list of available templates. This is the job submission script for LSF. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance or information about MPI, executable locations, etc. Perhaps the easiest method to find the correct LSF script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

For more details on optimization schemes, see Optimization Schemes.

Skip to next section.

1.4.5 Using a remote LoadLeveler cluster

lost:	somehost.com	SSH Port 22	0
Jser:	xtalopt-user		
Norking directory (Server):	/projects/thisrun		
Norking directory (Local):	/calcs/thatrun		
Description:	Imajob		
Path to Ilsubmit:	llsubmit		
Path to Ilcancel:	llcancel		
Path to Ilq:	llq		
Queue refresh interval:	20 sec		0
Clean remote directories	when finished		
		Cancel	Ж

Select "LoadLeveler" from the list of Queues, and then click the "Configure..." button. A new window will prompt for:

- · host: The hostname of the LoadLeveler cluster's head node
- · user: The username used to log into the cluster
- Working directory (Server): A directory that is readable/writable by "user" on the cluster, used when performing optimizations.
- Working directory (Local): A directory that is readable/writable by the current user on the local computer. This is where the final structures and resume files are written.
- · Description: Used for the description% keyword in input templates.
- Path to Ilsubmit: Where to find the Ilsubmit executable on the remote cluster. Note that if Ilsubmit is in the cluster's \$PATH, setting this to just 'llsubmit' will work.
- Path to llcancel: Where to find the llcancel executable on the remote cluster. Note that if llcancel is in the cluster's \$PATH, setting this to just 'llcancel' will work.
- Path to Ilq: Where to find the Ilq executable on the remote cluster. Note that if Ilq is in the cluster's \$PATH, setting this to just 'llq' will work.
- "Clean remote directories when finished": will remove all of the generated files from the cluster. Only the files on the local computer will be kept. If you do not want this to occur, make sure to uncheck this option.

A new template, "job.II" is added to the list of available templates. This is the job submission script for Load \leftarrow Leveler. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance or information about MPI, executable locations, etc. Perhaps the easiest method to find the correct LoadLeveler script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

For more details on optimization schemes, see Optimization Schemes.

Skip to next section.

1.4.6 Running optimizations locally

Local working di	rectory: /calcs/thatrun
Search descripti	on: Imajob
	Cancel OK

Select "Local" from the list of Queues, and then click the configure button. A new window will prompt for:

Local working directory: A directory that is readable/writable by the current user on the local computer. This
is where the final structures and resume files are written.

If the optimizer's executable (vasp, gulp, pw.x, castep, etc) is not in your system path, you will need to specify the location of the executable by clicking the "Configure..." button next to the optimizer selection menu.

For more details on optimization schemes, see Optimization Schemes.

1.5 What is written to the local directory?

A directory for each structure is created at

[Local working directory]/<gen#>x<id#>

that will contain input, output, and data files specific to each structure. Two additional files are also written to the local filesystem:

[Local working directory]/xtalopt.state

which contains save/resume information to continue a session that has been stopped, and

[Local working directory]/results.txt

which stores a list of all structures sorted by increasing enthalpy. The latter file is handy for offline analysis, since there is no need to open XtalOpt to find the most stable structures of a previous search.

1.6 Search Settings

	Structure Limits	Optimization Settings Sea	rch Settings Progress	Plot Lo	g
Initial Generation		Search Parameters		т	olerances
Initial structures: b0		Pool size per FL	: 20		Spacegroup perception
Initial Seed Structures:		Continuous structures	x 5		Length tolerance 0.050 Å
	_	Limit running jobs?	Limit to 1 jobs		Re-detect spacegroups
		If a job fails 2 times.	replace with random.		Duplicate matching
		Total Number of Structures	600		Length tolerance 0.100 Å
		Formula Unit Ontions			Angle tolerance 2.000°
Add	Remove	Allow FU Crossovers Use one gene pool? Chance of later mitosis: 50	r at Generation: 4 %	0	Reset duplicates
Crossover		Stripple		P	ermustrain
Percent new crossover: 15%	0	Percent new stripple: 50%	0		Percent new permustrain: 35%
Minimum contribution: 25%	0	Strain stdev range: 0.500	0.500 0		Maximum strain stdev: 0.500
		Amplitude range: 0.500	\$ 1.000		Number of exchanges: 4
		Waves in axis 1: 1	0		
		Waves in axis 2: 1	0		
ve Session Resume store	ed session	Optimized: 0	Running: 0 Fai	ilures: 0	Begin

In the "Search Settings" tab, most of the default settings should suffice (See CPC, 2011, 182, 372-387). We arbitrarily set the initial structures to 20 and the continuous structures to 5, although these may need to be adjusted based on available resources. We will not specify initial seeds, but the option to do so exists on this screen.

It is not necessary to limit the number of running jobs unless running locally, as the PBS queue on the cluster will manage job control for us. If running locally, set the job limit no higher than [number of available processor cores] - 1 (e.g. for a quadcore processor, allow three jobs to run simultaneously). This allows one core to remain free for the system to run.

There is now a termination criteria called "Total Number of Structures" that will end the run once a certain number of structures have been produced by XtalOpt.

The tolerances for duplicate matching are also found in this tab. They can be adjusted at any point in the run and the results.txt file will update with the correct duplicates found based upon the new tolerances.

Generated by Doxygen

1.7 "Begin"

	Sen	Mol	Job ID	Status	Time Flansed	Enthaloy (eV)	Volume	Space Group	Accestry
1 1	1	1	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	281.59	1: P1	Randomly generated
2 1	2	2	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	324.85	1: P1	Randomly generated
3 1	3	3	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	436.01	1: P1	Randomly generated
4 1	4	1	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	384.94	1: P1	Randomly generated
5 1	5	5	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	83.60	1: P1	Randomly generated
6 1	6	3	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	76.25	1: P1	Randomly generated
7 1	7	7	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	325.54	1: P1	Randomly generated
8 1	8	3	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	298.76	1: P1	Randomly generated
9 1	9)	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	226.71	1: P1	Randomly generated
10 1	1	10	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	359.45	1: P1	Randomly generated
11 1	1	11	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	246.84	1: P1	Randomly generated
12 1	1	12	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	245.57	1: P1	Randomly generated
13 1	1	13	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	238.33	1: P1	Randomly generated
14 1	1	14	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	492.63	1: P1	Randomly generated
15 1	1	15	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	437.93	1: P1	Randomly generated
16 1	1	16	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	156.03	1: P1	Randomly generated
17 1	1	17	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	493.89	1: P1	Randomly generated
18 1	1	18	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	193.47	1: P1	Randomly generated
Refre	esh	Refres	h period:	0 seconds				Print F	Results File Remove Extra Files Rank all Refresh al

Figure 1 The "Progress" tab immediately after starting a search

XtalOpt has everything it needs to start its search at this point; click the "Begin" button in the lower right corner of the application to tell it to start the search algorithm. A progress bar appears as the random first generation is created. Switch to the "Progress" tab and 20 entries will appear, all with a status of "Waiting for Optimization". Click "Refresh" on this tab to begin the local optimizations. From here, XtalOpt will continue to run without user input, starting new optimizations and generating new structures until it is stopped by the user.

1.8 Monitor progress

	Gen	Mol	Job ID	Status	Time Elapsed	Enthalpy (eV)	Volume	Space Group	Ancestry
20	1	20	N/A	Optimized	0:00:16	-230.119	286.73	1: P1	Randomly generated
21	2	1	N/A	Optimized	0:00:06	-238.802	182.20	136: P4_2/mnm	Stripple: 1x12 stdev=0.50000 amp=0.02264 waves=1,1
22	2	2	N/A	Duplicate of 1x2	0:00:04	-236.612	184.22	6: Pm	Stripple: 1x2 stdev=0.50000 amp=0.11003 waves=1,1
23	2	3	N/A	Optimized	0:00:11	-234.265	205.06	1: P1	Stripple: 1x17 stdev=0.22894 amp=0.53867 waves=1,1
24	2	4	N/A	Optimized	0:00:07	-235.109	174.88	6: Pm	Stripple: 1x12 stdev=0.18816 amp=0.97071 waves=1,1
25	3	1	N/A	Optimized	0:00:08	-233.592	185.11	4: P2_1	Stripple: 2x1 stdev=0.35948 amp=0.60497 waves=1,1
26	2	5	N/A	Duplicate of 1x12	0:00:20	-238.802	182.20	136: P4_2/mnm	Permustrain: 1x1 stdev=0.16277 exch=4
27	3	2	N/A	Optimized	0:00:26	-235.758	190.59	5: C2	Stripple: 2x1 stdev=0.50000 amp=0.49720 waves=1,1
28	3	3	N/A	Duplicate of 2x1	0:00:12	-238.802	182.20	136: P4_2/mnm	Stripple: 2x1 stdev=0.50000 amp=0.18255 waves=1,1
29	2	6	N/A	Optimized	0:00:04	-238.802	182.20	136: P4_2/mnm	Stripple: 1x12 stdev=0.12458 amp=0.75044 waves=1,1
30	2	7	N/A	Optimized	0:00:07	-234.446	182.16	6: Pm	Crossover: 1x2 (60%) + 1x19 (40%)
31	2	8	N/A	Optimized	0:00:14	-235.327	166.53	225: Fm-3m	Stripple: 1x12 stdev=0.41797 amp=0.51789 waves=1,1
32	4	1	N/A	Optimized	0:00:09	-238.59	180.75	14: P2_1/c	Stripple: 3x3 stdev=0.12166 amp=0.59499 waves=1,1
33	2	9	N/A	Optimized	0:00:14	-235.173	201.34	1: P1	Crossover: 1x19 (49%) + 1x4 (51%)
34	3	4	7419	Running (Opt Step 1 of 1, 0 failures)	0:00:25	N/A	182.20	1: P1	Permustrain: 2x6 stdev=0.38651 exch=4
35	4	2	N/A	Checking status	0:00:18	-232.786	214.89	1: P1	Stripple: 3x2 stdev=0.44728 amp=0.82453 waves=1,1
36	2	10	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	185.81	1: P1	Permustrain: 1x19 stdev=0.17759 exch=4
37	5	1	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	180.75	1: P1	Permustrain: 4x1 stdev=0.03405 exch=4
R	lefresh	Refre	sh period:	1 seconds		1		Print Resul	ts File Remove Extra Files Rank all Refresh all

Figure 2 The "Progress" tab mid-run

As XtalOpt performs the search, the progress table continuously updates, providing information about each structure. We see individuals in various stages of completion: most are optimized (in blue), structure 2x72 has been automatically marked as a duplicate (dark green) of structure 1x2 and removed from the breeding pool, structure 3x4 is currently undergoing a local optimization (light green), while structure 2x10 is waiting to be optimized (light blue).

Other useful information is displayed about each structure, such as the time spent in optimization, the optimized enthalpy, the cell volume, spacegroup, and each structure's ancestry (i.e. parent(s) and parameters for the genetic operator that generated it). A status bar on the bottom of the window shows the number of structures that are optimized, running, and failing at any given time. This information is visible regardless of which tab is currently being viewed.

After every structure has been optimized, it is checked to see if it is a super cell by performing a primitive reduction. If the primitive reduction yields a crystal with fewer atoms than it originally had, the original crystal is marked as a super cell and a new crystal is generated that is labelled as a primitive reduction of the original. This primitive-reduced crystal does not undergo optimization - it is labelled with the same energy as the original crystal. And, if the user is searching the formula units of the primitive-reduced crystal, the primitive-reduced crystal becomes a part of that smaller formula unit's gene pool.

In addition, every time a new crystal is optimized, if the user is searching multiple formula units and the crystal is found to be low in energy, a supercell of the crystal may be generated and be added to the higher formula unit's gene pool.

An additional feature of the progress table is the ability to immediately visualize any of the structures in the Avogadro2 main window (assuming Avogadro2 is open with an Avogadro2 RPC server running more info can be found in the Avogadro2 section) – simply clicking on a row in this table will display the threedimensional structure in Avogadro2, where it can be visualized, modified, or exported. If the user would like to add a bit of "intelligent design" to the evolutionary process, a structure can be modified and then resubmitted using a context (right-click) menu from the progress table. The context menu provides tools to (un)kill a structure, resubmit for local optimization at an arbitrary optimization step, replace a problematic structure with a new, random individual, or even generate a theoretical XRD pattern for the crystal.

Three additional buttons found near the "Refresh All" button in this tab are also available. The "Print Results File" button generates a run-results.txt file that lists all of the information about each structure in order of generation and structure number (As compared to the results.txt file which ranks the structures). The "Remove Extra Files" button is used for VASP runs. It removes any extraneous files in each local subdirectory in order to reduce disk usage. Files kept are structure.state, POTCAR, CONTCAR, OSZICAR, job.slurm and OUTCAR. Finally, the "Rank All" button ranks all currently optimized structures and exports them to a subdirectory (Ranked) in two forms (depending on the optimizer): POSCAR/.got and .cif. Each can be found in separate directories. (Only works for GULP and VASP runs currently).

To alter a run manually, right-clicking on any of the structures in the Progress tab will bring up a menu of options. These options include: Killing a structure, restarting a structure, replacing a structure, injecting a seed structure, etc. All of these can be done mid-run.

1.8.1 View trends



Figure 3 The "Plot" tab mid-run displaying enthalpy vs.\ volume. Each structure is labeled with its Hermann Mauguin spacegroup symbol.

Another visualization and analysis tool available during the search is the interactive plot. The plot is capable of investigating trends in the search by plotting a point for each individual using structure number, generation number, enthalpy, energy, \$PV\$ enthalpy term, lattice parameters, or cell volume on either axis. This powerful feature allows

the user to visualize complex relationships present in the generated structures. E.g., a plot of enthalpy vs. structure number provides an overview of the search's progress. Or, recalling that H = U + PV, plotting enthalpy vs. PV enthalpy term or energy lends insight into whether the enthalpy (H) is dominated by atomic interactions (U) or cell parameters (PV). Further information is available by labeling the points with the individual's spacegroup number, Hermann Mauguin spacegroup symbol, enthalpy, energy, PV term, volume, generation, or index number.

A particularly useful plot is that of enthalpy vs. cell volume, as shown above. From this view, we see a general trend that enthalpy increases with volume (the effect is much more pronounced for systems at higher pressures), and also that below a certain volume enthalpy rises sharply. From this data set, we see that there is a cluster of very low enthalpy structures with cell volumes around 180 cubic angstroms. Armed with this data, we can update the starting volume on the Cell Initialization tab mid-run to reflect this new piece of information that the search has provided us. Many of the other parameters governing structure generation and algorithm specifics can be similarly modified during a search without the need to restart the algorithm.

The plot is also interactive; zooming and panning are possible using simple mouse controls. Clicking on a structure's point on the plot will load it in the main Avogadro2 window (assuming Avogadro2 is open with an Avogadro2 R \leftarrow PC server running - more info in the Avogadro2 section), allowing all the same functionality as described above in Monitor progress.



1.8.2 View Crystals in Avogadro2

As of release 11, while an XtalOpt window is open, the user may easily view the crystals in Avogadro2. As long as an Avogadro2 window is open and an Avogadro2 Remote Procedure Call (RPC) server is running, when a user selects crystals in XtalOpt (either via the plot tab or the progress tab), the structure in the Avogadro2 window will automatically display the crystal the user selected. This allows for quick and easy visualization when analyzing a run.

1.8.3 Plot a Theoretical XRD Pattern

As of release 12, a theoretical XRD pattern can be generated for any crystal using Objcryst. A user must simply right click on any entry in the progress tab, and then click "Plot Theoretical XRD Pattern" (as shown in the image below).

Gen	Mol	Job ID	St	atus	Time Elapsed	H (eV/FU)	FU	Volume	Space Group	Ancestry
1	1	N/A	Waiting for Opt	imization (1 of 1)	0:00:02	N/A	2	57.75	47: Pmmm	RandSpg Init: 47 (P m m m)
1	2	N/A	Optimized		0:00:01	-39.3542	2	57.09	41: Aea2	RandSpg Init: 26 (P m c 21)
1	3	N/A	Optimized		0:00:01	-37.685	2	66.06	2: P-1	RandSpg Init: 2 (P -1)
1	4	N/A	Optimized	Pestart job	0.00-01	-30 8004	2	60.74	136: P4_2/mnm	RandSpg Init: 31 (P m n 21
1	5	N/A	Optimized	Kill structure			2	57.41	137: P4_2/nmc	RandSpg Init: 137 (P 42/n.
1	6	N/A	Optimized	– Reset <u>f</u> ailure	count		2	60.73	136: P4_2/mnm	RandSpg Init: 58 (P n n m)
1	7	N/A	Optimized	Replace with	new random stru	cture	2	60.73	136: P4_2/mnm	RandSpg Init: 16 (P 2 2 2)
1	8	N/A	Optimized	Replace with	new <u>o</u> ffspring		2	82.73	65: Cmmm	RandSpg Init: 65 (C m m m
1	9	N/A	Optimized	Inject <u>s</u> eed s	tructure		2	76.60	224: Pn-3m	RandSpg Init: 215 (P -4 3 n
0 1	10	N/A	Waiting for Opt	<u>C</u> opy POSCA	R to clipboard		2	67.53	13: P2/c	RandSpg Init: 13 (P 1 2/c 1
1 1	11	N/A	Waiting for Opt	Plot Theory	ical XRD Pattern	N/A	2	51.19	123: P4/mmm	RandSpg Init: 123 (P 4/m .
2 1	12	N/A	Optimized		0:00:01	-39.8004	2	60.73	136: P4_2/mnm	RandSpg Init: 100 (P 4 b m
3 1	13	N/A	Optimized		0:00:01	-38.6443	2	60.86	123: P4/mmm	RandSpg Init: 18 (P 21 21 2
4 1	14	N/A	Optimized		0:00:01	-38.3936	2	76.60	224: Pn-3m	RandSpg Init: 224 (P n 3 m
5 1	15	N/A	Optimized		0:00:01	-39.3046	2	57.40	137: P4_2/nmc	RandSpg Init: 5 (C 1 2 1)
6 1	16	N/A	Optimized		0:00:01	-36.1253	2	68.82	119: I-4m2	RandSpg Init: 119 (I -4 m 2
7 1	17	N/A	Waiting for Opt	imization (1 of 1)	0:00:00	N/A	2	58.51	119: I-4m2	RandSpg Init: 119 (I -4 m 2
8 1	18	N/A	Waiting for Opt	imization (1 of 1)	0:00:00	N/A	2	52.09	6: Pm	RandSpg Init: 6 (P 1 m 1)
~ 4	40	11/4			0.00.00	N1/A	<u> </u>	CA CO	04:04	D dC 1-11- 04 (D4)

A dialog box will appear that allows a user to select several options for the XRD pattern. All of these options have tooltips that describe their function. The options include "Wavelength" (the wavelength of the x-ray), "Peak width" (which broadens the peaks to help simulate temperature), "Number of points" (to increase smoothness), and "Max 2*theta" (the max value to be displayed on the x-axis).

Generate Theoreti	cal XRD Pattern Opt⊗
Wavelength:	1.50560 Å
Dealswidth	0.52050
Peak width:	0.52958
Number of points:	1000
Max 2*theta:	162.00°
	≭ <u>C</u> ancel

Once these options have been chosen, the user may simply click "OK" to generate the plot. A plot will appear similar to the one shown below.



1.9 Command Line Interface

As of release 11, there is a command line interface (CLI) available within XtalOpt. This allows users to choose settings and run searches without the use of the GUI.

First of all, one can see all of the XtalOpt CLI options by running "xtalopt --help" from a terminal. This prints all options and a description of each one.

When performing a CLI run, an input file for XtalOpt is required. An example which contains detailed explanations of various options can be found here.

A CLI run is begun by entering into the terminal "xtalopt --cli". By default, XtalOpt will search for an "xtalopt.in" file in the current directory to use for the input file. If a different input file is desired, the user may explicitly select an input file with "--input-file <file>". The vast majority of the options are taken from the GUI, and more information on these options can be be found in their respective places in the manual. One difference is that template files need to

be created to for the different templates for queue interfaces and optimization schemes. These templates use the same %keywords% that are used in the GUI, and an example of a GULP template can be found here.

When a CLI run first begins, all of the options set by the user and XtalOpt's resulting settings will be printed to the terminal (and to an "xtaloptSettings.log" file in the local working directory). The user should glance over these printed options to ensure that they are set correctly. While a CLI run is in progress, the user is still able to update settings via a file in their local working directory called "xtalopt-runtime-options.txt". The settings in the "xtalopt-runtime-options.txt" file are read every iteration of the event loop. Thus, if a setting is changed in that file, it will quickly be updated in the program.

In addition, while the CLI run is in progress, job submissions, completions, and errors will be reported in the terminal. If one wishes to end the CLI run at any time, they just need to use "ctrl-C" or whatever their process interruption command is.

One is also able to resume a run via the CLI if they type in "xtalopt --resume --dir <path/to/resume/dir>". This will check to see if an "xtalopt.state" file is in the specified directory, and if it is, XtalOpt will attempt to resume the run.

While a CLI run is in progress (or after any run is finished), the user may view the enthalpy/energy results via the "results.txt" file in the local working directory. They may also generate a plot with the "xtalopt --plot --dir <path/to/dir>" command. This will immediately display a plot using the same code as that used to generate the plot tab in a GUI run. Similar to a GUI run, plot axes and other options may be changed. In addition, if Avogadro2 is open in the background and an Avogadro2 RPC server is running, XtalOpt will still also set the structure in the Avogadro2 view to the crystals the user selects.

A note for Linux users: if a window manager is not running on the computer attempting to use CLI mode, the program may still crash. One solution to get around this is to add the "-platform offscreen" option to the command line arguments (this will only work if X11 is installed on the computer). This requirement will hopefully be removed in the next XtalOpt update.

		Optimization Settings	Search Settings	Progress Plot I	_og			
Composition			Unit Cell Paramet	ers				
Cell composition:				Minimum		Maximum		
Symbol Z # Mass	Min. Radius		Length A (Å):	1.00000		20.00000		
			Length B (Å):	1.00000		20.00000		
			Length C (Å):	1.00000		20.00000		
			Angle α (°):	60.00000		120.00000		
			Angle β (°):	60.00000		120.00000		
			Angle y (°):	60.00000		120.00000		
			Volume (Aª):	1.00		500.00		
			E Fiz	xed volume Vc	lume (Aª):		10	
			Mi	itosis	# of	Divisions:		
			Ex	port subcells? a:				
			🗾 Lir	mit interatomic distance	9 8	cale factor:	0.50 * radii	
					Minim	um radius:	0.25 Å	
ave Session Resume stored session		Optimized:	0 Running:	0 Failures: 0				Begin

2 Saving and Resuming Sessions in XtalOpt

2.1 How to save your session

XtalOpt will write a small file named xtalopt.state to its working directory that contains the information necessary to resume the session at a later time. The file can be rewritten manually by clicking the "Save Session" button highlighted above, and XtalOpt will automatically save the session every time a structure is updated.

XtalOpt will also write a file "structure.state" in each candidate structure's directory. This file stores XtalOpt-specific information about the structure.

2.2 How to resume your session

To resume a session, simply click "Resume stored session" (highlighted above) and select the xtalopt.state file in the working directory of the session you would like to resume. XtalOpt will then begin to load the structures and search parameters. You can monitor the progress with the progress bar that appears at the bottom of the window.

While the structures are loading, you may encounter errors that say:

Error, no (or not appropriate for [OPTIMIZER]) xtal data in [DIRECTORY]. This could be a result of resuming a structure that has not yet done any local optimizations. If so, safely ignore this message.

As mentioned in the message, these can typically be ignored if it only happens for a handful of structures. This occurs when a structure has been generated in XtalOpt, but it has not completed any geometry optimization so there are no output files from which to load the geometry. If it happens for a significant number of structures (or structures that are known to have completed at least one geometry optimization step), the output files from the optimizer may be missing or corrupt.

After resuming a session, XtalOpt will ask if you would like to continue the search or enter read-only mode. Readonly mode will not generate new structures or submit geometry optimizations.

Note

If you are considering resuming a read-only session, take a look at the results.txt file in the working directory. It contains a list of all structures, sorted by enthalpy, with additional useful information. This can save some time when trying to locate the most stable structure of an old search.

The working directories for XtalOpt are relocatable, meaning that the directory containing xtalopt.state and the [gen]x[id] structure folders may be moved, tarred, zipped, etc. and still be resumed at a later time from a different location on the filesystem, or even a different computer.

3 Optimization Schemes

- 3.1 Overview: What are optimization schemes, and why use them?
- 3.1.1 In a nutshell..

An optimization scheme is a series of optimization steps ("optsteps") that are to be performed in sequence on a structure. Each optimization step consists of a set of input file templates for the queuing system and optimizer to be used, and the structure is updated after each completes. So if an optimization scheme contains three optimization steps, a structure's lifecycle is:

- 1. Generation of initial structure
- 2. Perform optstep 1 on initial structure
- 3. Update structure from the results of optstep 1
- 4. Perform optstep 2 on current structure (result of optstep 1)
- 5. Update structure from the results of optstep 2
- 6. Perform optstep 3 on current structure (result of optstep 2)
- 7. Update structure from the results of optstep 3
- 8. Current structure (result of optstep 3) is either accepted into the breeding pool or discarded, depending on its enthalpy relative to the other optimized structures.

3.1.2 More details

The efficiency of searching a potential energy surface for a global minimum can be significantly improved by moving each candidate structure to the nearest local minimum, i.e. performing a geometry optimization. The differences between searching with and without carrying out these local optimizations are explored in detail in Woodley SM, Catlow CRA. Comp. Mat. Sci. 2009;45(1):84-95 (Available at: http://linkinghub.elsevier.↔ com/retrieve/pii/S0927025608003030).

Why not just perform a single geometry optimization on each structure? Stochastic search techniques, such as XtalOpt, will often need to perform geometry optimizations on structures that are far from a stationary point on the potential energy surface. For example, the randomly generated structures in the first generation of an evolutionary search are often highly disordered with unrealistic atomic separations. If these structures were to be optimized in a single step with accurately small convergence criteria, it would be quite expensive. Also, it is more than likely that most of the optimizations would not finish successfully before reaching the maximum number of geometry steps allowed by the optimizer or specified in the input. A second issue is that complex structures (periodic crystals, for example) often have so many degrees of freedom that convergence in a single step is difficult from a poor starting point (consider the effect on atomic coordinates when a unit cell's translation vector is modified).

The first problem (effectively optimizing to small convergence) can be solved by implementing an optimization scheme that optimizes to successively smaller convergence cutoffs.

The second problem can be addressed by reducing the degrees of freedom in the early optsteps and only optimizing everything once each component has individually converged to a reasonable parameterization. See Suggestions for optimization schemes for examples.

3.2 Optimization scheme user interface

		TtalOpt	
		Structure Limits Optimization Settings Search Settings Progress Plot Log	
Queue PBS	Configure	\$filename\$	
Optimizer VASP	Configure	<pre># output options IMAVE = .FALSE. # write or don't write WAVECAR ICHARG = .FALSE. # write or don't write CRG and CHGCAR</pre>	
Template INCAR	40	LELF - FALSE. # write ELF	
Optimization 1 Optimization 2		# ionic relaxation NSW = 20	
Optimization 4 Optimization 5 Optimization 6		# provision parameters EDTT = 1.2	
		f electronic relaxation ISMRAB = 5 f + 5 = tetraedon, 1N = Methfessel ISMCT = %user3 f outoff energy FerRES = %user24	
		ISYR=U	
	1		
0	ove		
Add Rem			
Add 2 Rem	Load Opt Scheme 7		
Add. 2 Rem	Load Opt Scheme		
Add 2 Rem Save Opt Scheme	Load Opt Scheme		
Add 2 Rem Save Opt Scheme user1: user2: 0	Load Opt Scheme		
Add Rem Save Opt Scheme user1: user2: 0 user3: 400	Load Opt Scheme		5
Add. Rem Save Opt Scheme user1: user2: 0 user3: 400 user4:	Load Opt Scheme		5
Add Rem Save Opt Scheme user1: user2: 0 user3: 400 user4:	Load Opt Scheme		5

We will use the above screenshot as we describe the process of creating, saving, and loading optimization schemes. The numbers indicate:

- 1. List of optimization steps
- 2. Button to add new optimization step
- 3. Button to remove current optimization step
- 4. Template selection menu
- 5. Template editor
- 6. Button to save current optimization scheme to file
- 7. Button to load optimization scheme from file

3.2.1 Optimization step list

This list shows the currently available optimization steps in the order that they will be performed. The optstep that is currently selected for editing is highlighted, and the editable optstep can be selected by clicking the appropriate entry.

3.2.2 Add new optimization step

Clicking this button will append a new optimization step to the optstep list. The new optstep's templates will be copies of the currently selected optstep's templates.

3.2.3 Remove current optimization step

Click this button to delete the currently selected optimization step.

3.2.4 Select template

This menu contains the filenames of the templates that are required by the currently selected queuing system (e.g. PBS, SGE, local...) and optimizer. The currently selected template is displayed in the template editor, and selecting a different template will update the editor.

3.2.5 Template editor

This text editor is used to view and edit the currently selected template for the current optstep.

3.2.6 Save scheme

This button will prompt for a location to save a .scheme file containing the current optimization step.

3.2.7 Resume scheme

This button will prompt for an existing .scheme file to load.

3.3 How to build an optimization scheme?

Creating a working scheme from scratch may take some time. We recommend checking the samples/ directory of the source code to obtain sample scheme for each optimizer (see How to load an optimization scheme?) and verifying that they are appropriate for the system under consideration before starting a search.

If there is not an appropriate sample, the following prescription may be used to generate your own:

- 1. Generate a random structure of the system under consideration. This may be done by hand, or by running a search just long enough to create the first random generation and saving one of the structures.
- 2. Create a starting optstep with the desired convergence criteria
- 3. Manually submit the optimization
- 4. If the optimization fails:
 - (a) First determine why if the maximum iterations were exceeded or the optimization was aborted due to a badly performing minimizer, try one of the ideas below. Other optimization problems are beyond the scope of this document.
 - (b) Reduce the convergence criteria of the current trial optstep
 - (c) Remove degrees of freedom, e.g. by fixing cell parameters, atomic positions, etc.
 - (d) Reduce the accuracy of the calculation in other ways (use a courser integration grid, etc).
 - (e) Change the minimizer (e.g. tell the optimizer to use conjugate gradients rather than BFGS, etc)
- 5. Once the optimization succeeds, create another set of input files with the desired convergence criteria for all degrees of freedom.

- 6. Manually submit the new optimization step. If it fails, try the ideas above until it converges.
- 7. Once the structure has converged to the desired level of accuracy, try to optimize another randomly generated structure using the optsteps that succeeded previously. Refine them if needed.
- 8. Once you have successfully optimized enough random structures that you are confident in your method, gather all of the inputs used and write your scheme from them.

The scheme may be written by copying each input file into the template editor (with the appropriate optstep and template selected, of course) and replacing the structure-specific information with the appropriate keywords. Click the "Help" button for the complete list of keywords.

We have found that the optimization schemes are surprisingly transferable within an optimizer, so once you have a working optimization scheme for a given optimization code only minor tweaks (usually to the energy cutoffs, etc.) are necessary to use it on a different chemical system.

It is important to note that the optimization scheme does not have to perfectly converge your structures. A final post-processing optimization to refine any structure found in the search is highly recommended.

3.4 How to save an optimization scheme for later?

Once you have written your optimization scheme, you will want to save it for fast retrieval later (otherwise you will need to copy/paste and edit all of the templates again!). To save, simply click the "Save Opt Scheme" button and enter an appropriate filename with an extension of .scheme.

3.5 How to load an optimization scheme?

Loading an optimization is quite simple – just click the "Load Opt Scheme" button and select the .scheme file you wish to load. This will also update the current queuing system and optimizer to those specified by the scheme.

3.6 What is saved?

The optimization scheme files contain more than just the templates for each optstep. They also store queue and optimizer specific settings. This is useful for storing configuration options for different clusters along with the scheme. Note that although XtalOpt will prompt for an SSH password if needed, it is **NOT** stored in the scheme file.

3.7 Suggestions for optimization schemes

3.7.1 Crystals (XtalOpt)

The following list describes the optimization steps used in the samples/vasp-xtalopt.scheme file distributed with the XtalOpt source code:

- 1. Fix unit cell, only optimize atomic coordinates. A very loose convergence criterion is used, and the number of KPOINTs is kept small.
- 2. The cell volume is fixed, but atomic positions and cell parameters are allowed to vary. The convergence criteria is the same as before, as is the KPOINT grid.
- 3. All degrees of freedom are considered using the same convergence criteria as before, but with a finer KPOINT grid.
- 4. Same as before, but with a stricter convergence criteria.
- 5. Same as before, but with a stricter convergence criteria and more KPOINTs.
- 6. Same as before, but with more KPOINTs.

This is only one of many possible optimization schemes that may work for crystals. It may need to be modified to work for your particular system.