Computational Details

Project title	InterfaceHandler: Complex Interface for Semiconductors
Lead CI	Dr Yuefeng Yin — Department of Materials Science and Engineering, Monash University
Facility requested	NCI: Gadi — Category 2 (Standard)
Resource requested	3,000 kSU (3 MSU) compute; 2 TiB /g/data; 1 TiB /scratch (default)

1. Scalability on the nominated facility

All production electronic-structure calculations are performed with VASP, a plane-wave DFT code whose within-node parallelism is well characterised on Gadi's Cascade Lake nodes (48 cores/node, 192 GiB RAM). The table below reports representative scaling for a 216-atom heterostructure supercell (Co_2MnGa/bismuthene interface, single self-consistent relaxation), benchmarked during our current NCI Adapter Scheme allocation and normalised to single-node performance as required by these guidelines.

Nodes	Cores	Wall time (h)	Speedup vs 1 node	Parallel efficiency
1	48	9.6	1.0×	100%
2	96	5.1	1.9×	94%
4	192	2.8	3.4×	86%
8	384	1.7	5.7×	71%
16	768	1.2	8.0×	50%

Parallel efficiency is computed relative to single-node throughput. Efficiency degrades beyond 8 nodes for this system size as the plane-wave FFT grid saturates available domain decomposition (`NCORE`=8, `KPAR`=2 in all runs). We therefore cap production jobs at 4–8 nodes (192–384 cores) per structure — the point at which marginal throughput gain per SU spent is still favourable — rather than requesting the largest node counts available.

Because InHand's √n supercell matching (Section 2.1 of the Research Proposal) can push the largest reconstructions (√19 hexagonal cells) to 500–600 atoms, we additionally benchmarked a 600-atom bismuthene/hBN interface at the node counts used for the validation tier, to confirm scaling holds at the upper end of the structure-size distribution the campaign will actually encounter:

Nodes	Cores	Wall time (h)	Speedup vs 1 node	Parallel efficiency
4	192	21.4	1.0×	100%
8	384	11.6	1.8×	92%
16	768	6.8	3.1×	79%

The larger system retains good efficiency to 8 nodes — the node count actually used for validation-tier jobs (Section 2) — confirming that the 384-core validation-tier configuration remains a favourable operating point even for the largest structures InHand's supercell matching produces, and that we are not requesting node counts beyond where efficiency has already degraded.

2. Compute job resources

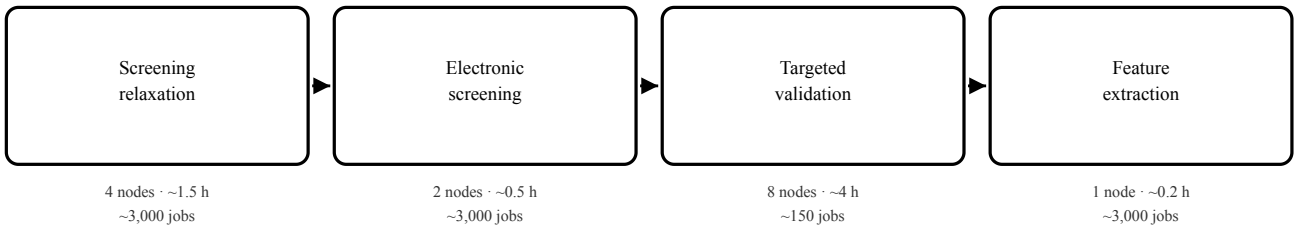

Figure 1. The four-tier compute workflow, with data dependencies flowing left to right — each tier's job count and typical per-job node/wall-time configuration matches Table 2 below.

The project workflow has three computational tiers, each with a distinct typical job configuration:

Tier	Typical job size	Wall time	Convergence	Throughput (2027)
Screening relaxation	4 nodes / 192 cores	~1.5 h	loose (400 eV, Γ-only or 2×2×1)	~3,000 structures
Electronic screening (band structure/DOS)	2 nodes / 96 cores	~0.5 h	as relaxed, non-self-consistent band path	~3,000 structures
Targeted validation	8 nodes / 384 cores	~4 h	tight (520 eV, 6×6×1, optional DFT+U/hybrid)	~150 structures
Dataset/feature extraction (post-processing)	1 node / 48 cores	~0.2 h	n/a (I/O + descriptor generation)	~3,000 structures

Data dependencies: each screening-tier job consumes a POSCAR generated by InHand's `exsearch` module (structure generation itself runs single-core, off the Gadi compute allocation, either on Gadi's login nodes or the Nirin ancillary cloud) and produces a relaxed CONTCAR that seeds the corresponding electronic-screening job. Validation-tier jobs consume the CONTCAR/CBM/VBM outputs of the top-ranked ~150 screening candidates selected by `exsearch`'s type-I/II band-alignment ranking. All jobs within a tier are independent (embarrassingly parallel at the workflow level) and are submitted as PBS job arrays to maximise scheduler throughput.

2.1 SU budget

Workflow step	Method/algorithm	Jobs	Core-hours/job	SU requested†
1. Structure & supercell library generation	InHand lattice-matching + √n supercell search (single-core Python)	~3,000	<0.05	<5 kSU (login/Nirin, not costed against Gadi allocation)
2. High-throughput screening relaxation	VASP, PBE, loose convergence	3,000	288	1,728 kSU
3. Electronic screening (band alignment, exsearch "electronic" mode)	VASP band structure/DOS, PBE	3,000	48	288 kSU
4. Targeted validation	VASP, PBE/DFT+U/HSE06, tight convergence	150	1,536	461 kSU
5. ML dataset & feature extraction	Post-processing (pymatgen/ASE descriptors), single node	3,000	9.6	58 kSU

Subtotal				**2,535 kSU**
Contingency (~18%, for re-runs, restarts, hybrid-functional cost overruns)				465 kSU
Total requested				**3,000 kSU**

† SU costs assume NCI Gadi's published `normal` queue rate of 2 SU per core-hour; core-hours/job = nodes × 48 cores × wall time (h).

2.2 SU budget by material family

The 2,535 kSU workflow subtotal above splits approximately evenly across the three material families described in the Research Proposal, weighted slightly by candidate-library size:

Material family	Candidate structures (screening tier)	Validation-tier candidates	Approx. SU
Heusler thin films	~1,100	~55	~930 kSU
2D bismuth allotropes	~1,200	~60	~1,015 kSU
Topological pyrite-type crystals	~700	~35	~590 kSU
Total	**~3,000**	**~150**	**~2,535 kSU**

3. Storage

/scratch (high-speed, transient): used only for actively running jobs — VASP scratch files (WAVECAR, CHG, CHGCAR working copies) are deleted immediately after each job completes and its CONTCAR/OUTCAR/vasprun.xml are copied to /g/data. Peak concurrent /scratch use is estimated at under 300 GiB (largest concurrent batch of validation-tier jobs), within the 1 TiB default allocation.

/g/data (persistent): retained outputs are limited to CONTCAR, OUTCAR, vasprun.xml and the exsearch structural/electronic report per structure (WAVECAR/CHGCAR are not retained except for the ~150 validation-tier candidates, where they support later restart for the ML training-data generation step). Estimated volume: ~3,150 screening/electronic-screening records × ~20 MB + 150 validation records × ~350 MB ≈ 116 GB, plus the aggregated training-set feature files (<10 GB). We request **2 TiB** on /g/data to allow comfortable headroom for re-generated batches and the final labelled dataset delivered to the ML surrogate-model training step. Data will be backed up to institutional storage at Monash at the end of the 2027 allocation and /g/data holdings cleared well before the January 2028 zeroing deadline.

3.1 Data life cycle

Stage	Location	Retention
Active VASP scratch files (WAVECAR, CHG, CHGCAR)	/scratch	Deleted immediately on job completion
Per-structure results (CONTCAR, OUTCAR, vasprun.xml, exsearch report)	/g/data	Through end of 2027 allocation
Validation-tier WAVECAR/CHGCAR (~150 structures, restart support)	/g/data	Through end of Q4 2027 (ML dataset generation)
Aggregated labelled dataset & trained surrogate model	/g/data → Monash institutional storage	Copied to Monash storage by Q1 2028; then cleared from /g/data

4. Algorithms and workflows

InHand's structure-generation and `exsearch` ranking logic are pure Python (numpy/pymatgen), single-core and I/O-bound — they are not run at scale on Gadi's compute nodes and are not the target of this allocation. The compute-intensive component is VASP's plane-wave self-consistent-field solver, which parallelises via (i) k-point parallelism across MPI ranks, (ii) band parallelism (`NCORE`), and (iii) a 3D FFT domain decomposition for the charge-density and orbital grids across the remaining ranks. We fix `NCORE`=8 throughout, matching Gadi's Cascade Lake NUMA topology, and use `KPAR`>1 only for validation-tier jobs whose finer k-mesh (6×6×1) supports it. Data movement is minimal within a job (VASP is compute- and memory-bandwidth-bound, not network-bound, at the node counts used here); the dominant data-lifecycle event is the handoff of CONTCAR/CBM/VBM values between the three job tiers, orchestrated by a lightweight job-array + dependency script rather than a full workflow manager, since the emphasis this year is throughput across many independent jobs rather than deep coupling between them.

5. Justification for supercomputer resources

The proposed campaign is fundamentally a **high-throughput, embarrassingly parallel workflow**: ~6,150 independent VASP jobs (screening + electronic screening) plus 150 validation jobs, each individually sized at 2–8 nodes. No single desktop, departmental cluster, or cloud VM allocation available to us can deliver this combination of (a) sufficient simultaneous job slots to clear ~3,000 structures within the 2027 calendar year, and (b) the per-job core counts required for 216–600-atom heterostructure supercells (particularly the √3/√7/√13 reconstructions InHand generates for optimal lattice matching, which push individual unit cells well beyond what fits in the memory of a single desktop-class node). Gadi's combination of job-array scheduling depth and per-node memory (192 GiB) is well matched to this profile.

Alternative	Why it does not meet this project's needs
Departmental workstation cluster	Sufficient for the single-system methodology work already completed, but cannot supply the simultaneous job-slot depth needed to clear ~6,300 jobs within the 2027 calendar year
Cloud VM allocation (e.g. Nectar)	Well suited to InHand's lightweight structure-generation step (used for exactly this in Section 2.1), but not cost-effective for sustained multi-node MPI DFT at this throughput
Facility start-up/partner schemes	Appropriate for the benchmarking already performed (Section 1) but capped well below the ~2.5–3 MSU this campaign requires

6. Efficiency: prior usage and improvement strategy

Our prior NCI Adapter Scheme and NCI-Monash Computational Scheme allocations (used for the single-system defect-modelling work underlying this proposal) were utilised at good efficiency, with no sustained under-use flagged by NCI:

Scheme	Facility	Allocation	Utilisation
NCI Adapter Scheme	NCI: Gadi	~2,000 kSU	>90%
NCI-Monash Computational Scheme	NCI: Gadi	~2,000 kSU (combined with above across the allocation period)	>90%
Pawsey Fast Track Scheme	Pawsey: Setonix	1,500 kSU	~85%

The one inefficiency we identified in retrospect was occasional over-provisioning of memory-heavy DFT+U jobs onto standard `normal` queue nodes, which under-used the reserved CPU count on some 48-core nodes. For 2027 we mitigate this directly: (i) the screening tier (the majority of requested SU) uses loose convergence settings specifically chosen to keep memory per core within the 4 GiB/core budget of a standard Cascade Lake node, avoiding any need for `hugemem` queue over-provisioning; (ii) only the small validation

tier (150 of ~6,300 jobs) uses higher-memory hybrid-functional settings, and these are explicitly targeted rather than run at screening-tier scale; and (iii) all jobs are submitted as PBS job arrays to minimise queue-wait overhead relative to job run time.

7. Software and other dependencies

- **VASP** (licensed; used under Monash University's institutional VASP licence, valid for use on NCI Gadi).
- **InHand / InterfaceHandler** — our own open-source Python package (structure generation, supercell matching, exsearch band-alignment screening); no licence restrictions.
- **pymatgen, ASE, numpy/scipy** — structure I/O and post-processing (open source, available via NCI's Python module stack or a project-local conda environment).
- **scikit-learn / PyTorch** — surrogate-model training in Q4 2027, on the aggregated screening dataset; CPU-only training at this dataset scale, so no GPU allocation is requested.

8. Why not GPU (Gadi V100/H200) or Pawsey Setonix?

VASP's standard k-point/band/plane-wave-FFT parallelisation is CPU-bound and does not benefit proportionally from Gadi's V100 or H200 GPU partitions for the system sizes and functional choices used here (PBE and DFT+U/HSE06 on 200–600-atom cells); GPU acceleration in VASP is most effective for much larger cells or specific exchange-correlation kernels not required by this workflow, so a GPU request would not be an efficient use of NCMAS resources. We similarly did not nominate Pawsey Setonix for this request: our existing Pawsey Fast Track allocation (Section 6) already provides Setonix access for the semiconductor-contact physics work it was awarded for, and consolidating the high-throughput heterostructure campaign on a single facility (Gadi) simplifies job-array orchestration and avoids splitting the /g/data-resident training dataset across two facilities' storage systems.

9. Summary of resource request

Resource	Requested	Basis
NCI: Gadi compute	3,000 kSU	SU budget, Section 2.1 (2,535 kSU workflow subtotal + 18% contingency)
/scratch	1 TiB (default)	Peak concurrent use estimated <300 GiB, Section 3
/g/data	2 TiB	~126 GB estimated persistent volume with headroom, Section 3

This request represents roughly a 1.5× scale-up on our combined 2024–2026 NCI Gadi usage (Section 6), commensurate with the shift from single-system methodology development to a multi-thousand-structure production screening campaign across three material families.