WP3: Computing Requirements

Anna Scaife, Mark Ashdown

Objectives

- WP3 will identify and assess the components necessary to bring about a European Science Data Centre, both in hardware and software, from a **total science delivery perspective**
- The focal questions are:
  - *"What does the ESDC need to do to maximise European science delivery from the SKA?"*
  - *"How can we build such a science data centre, and at what cost?"*
- WP3 will develop a set of **recommendations** for the design of the ESDC pertinent to its data handling strategy, scientific functionality and software environment
- WP3 will produce a **high level architectural design** for the ESDC and a **sizing and costing** estimate

Deliverables

- **D3.1** Analysis of compute load, data transfer and data storage anticipated as required for SKA Key science (M21)
- **D3.2** Report on suggested solutions to address each of the key software areas associated with running a distributed ESDC (M24)
- **D3.3** Preliminary System sizing report (M24)
- **D3.4** Report on design & costing for ESDC (M36)
- **D3.5** Report on suggested solutions to interface requirements for a distributed ESDC (M36)

D3.1 Analysis of Computing and Storage Requirements

- Based on an analysis of the 13 **High-Priority Science Objectives (HPSOs)**
- These were selected by SKA, based on inputs from the **Science Working Groups**
- Taken to be representative of the future **Key Science Projects**
- HPSO observations finish after some period of time
  - In a real observing programme, new projects will come along to replace them

Use Cases

Use cases were selected:

1. to be representative of a wide range of processing models;
2. to cover a range of SKA science working groups;
3. to use a variety of SDP data products as input data;
4. to be high usage cases, i.e. they will need to be run as standard processing on the majority of datasets.

Table 1: List of SKA Science Working Groups.

SKA Science Working Groups			
1	Extragalactic Spectral Line	7	Our Galaxy
2	Solar, Heliospheric & Ionospheric Physics	8	Epoch of Reionization (EOR)
3	Cosmology	9	Extragalactic Continuum
4	Cradle of Life	10	HI Galaxy Science
5	Magnetism	11	Pulsars
6	Transients		

Table 2: Summary table of processing Use Cases used within WP3.

No.	Name	Input Data	SWGs
1	Calibration & Imaging	Calibrated Visibilities	**1**, 3
2	Pulsar Re-folding	Pulsar Candidates	**11**
3	Rotation Measure Synthesis	Image Cube [4]	**5**
4	Object Detection and Classification	Image Cube [1]	1, 3, 4, 5, 6, 7, **9**, 10
5	Automated Object Classification	LSM Catalogue	1, 3, 4, 5, **6**, 7, 9, 10

Data Storage Requirements

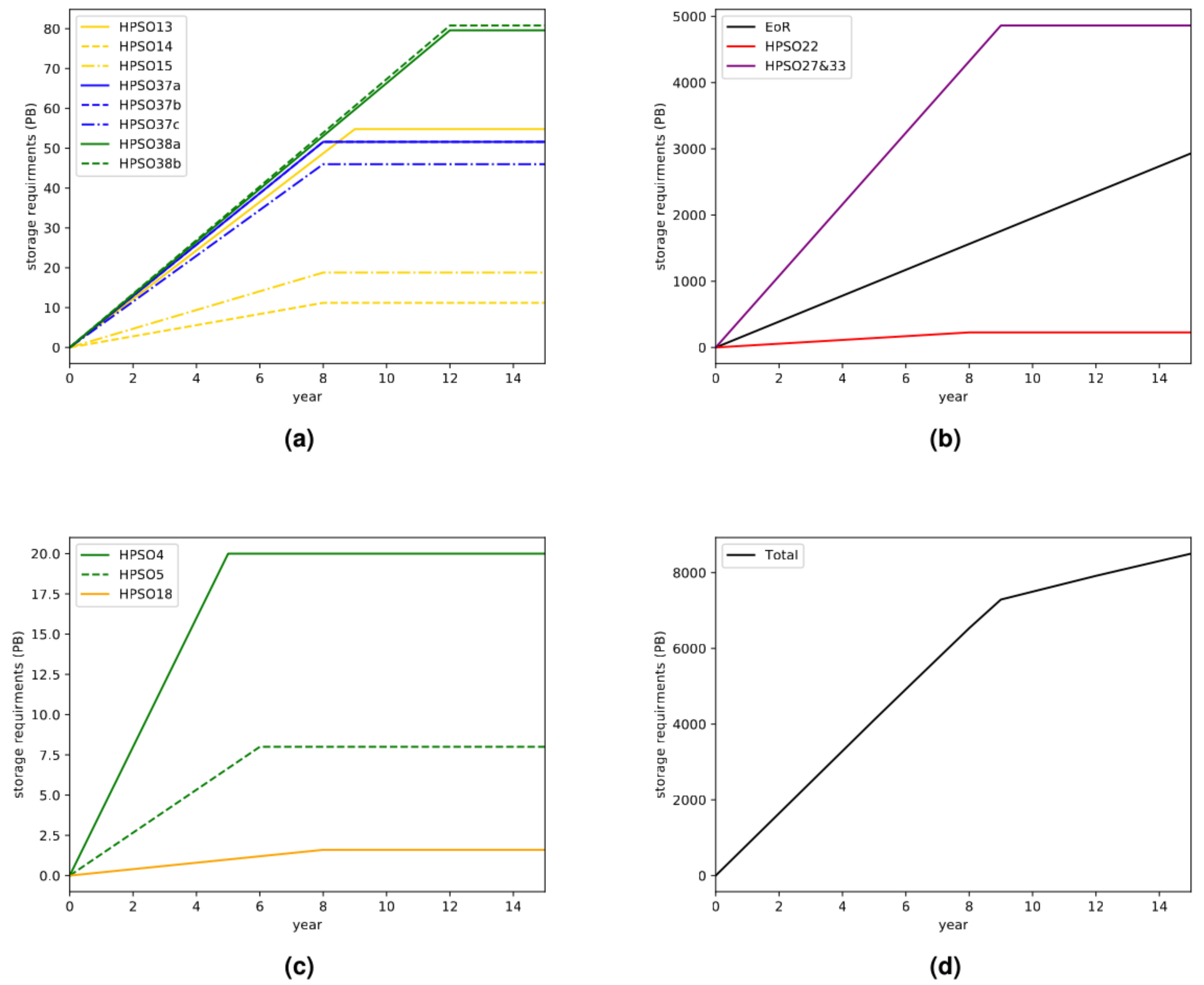

Figure 1: (a) Data storage requirements at SRCs for the HI and continuum HPSO. (b) Data storage requirements at SRCs for the EoR, magnetism and cradle of life HPSOs. (c) Data storage requirements at SRCs for the pulsars and transients HPSOs. (d) Data storage requirements at SRCs for all HPSOs.

Storage **volume** requirements depend on:

- Rate of ingest of Observatory data products;
- Rate of production of Advanced data products

Rate of ingest of Observatory data products was based on the SDP System Sizing, assuming that the ESRC holds a complete copy of all products

Rate of production of advanced data products was determined using the AENEAS processing Use Cases and was determined to be **3:1 in volume** (output:input)

Also examined the **number** of data products expected as a function of time. Relevant for the choice of Data Management System (DMS)

Data Storage Requirements

Minimum storage needed to meet the needs of the HPSOs:

- **744 PB per year for the first 10 years of operation**, and
- **201 PB per year for the following 5 years**

The HPSO storage volume required by the SRC network is therefore **8.5 Exabytes over the course of 15 years**

Storage **volume** requirements depend on:

- Rate of ingest of Observatory data products;
- Rate of production of Advanced data products

Rate of ingest of Observatory data products was based on the SDP System Sizing, assuming that the ESRC holds a complete copy of all products

Rate of production of advanced data products was determined using the AENEAS processing Use Cases and was determined to be **3:1 in volume** (output:input)

Also examined the **number** of data products expected as a function of time. Relevant for the choice of Data Management System (DMS)

Data Processing Requirements

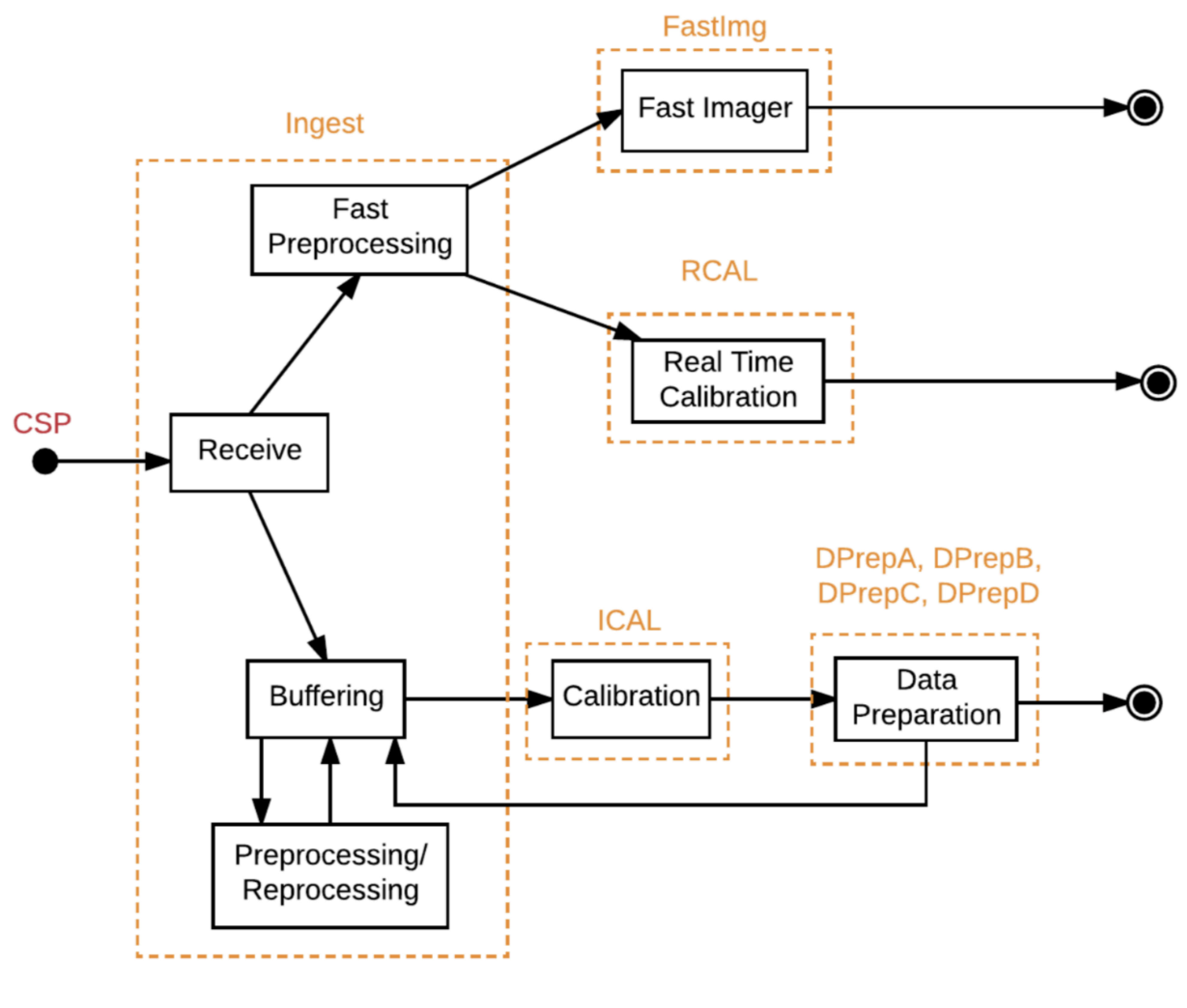

Figure 1: Overview of SDP pipelines [A07].

Processing was examined in terms of

1. compute load;
2. memory requirements;
3. potential for distribution;
4. suitability of platform.

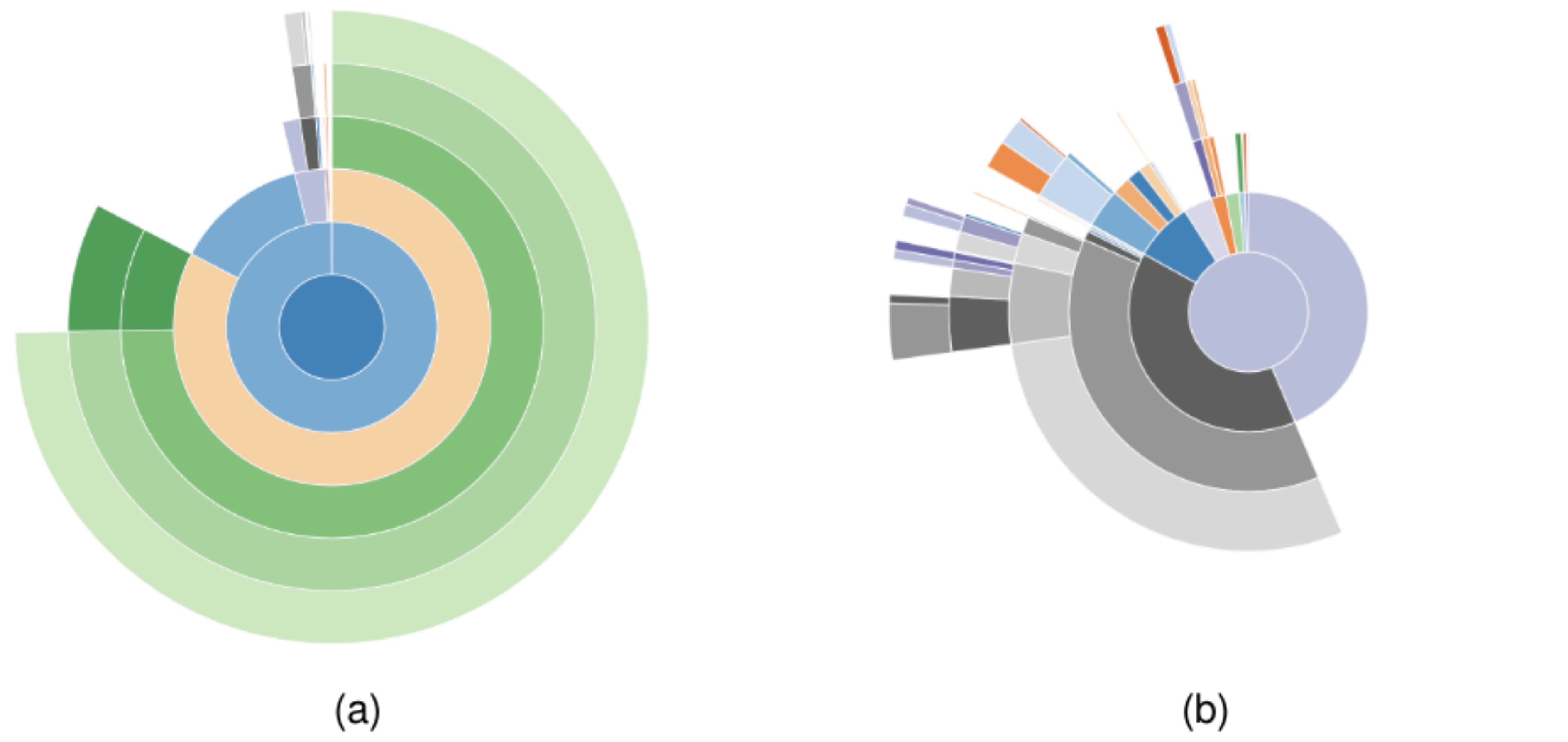

Figure 9: Visualisation of processing time breakdown for PyRMSynth. (a) The overall breakdown, dominated by `write_output_files` (green); (b) the breakdown beneath `compute_dirty_image`, dominated by the FFT (grey). See text for details.

Data Processing Requirements

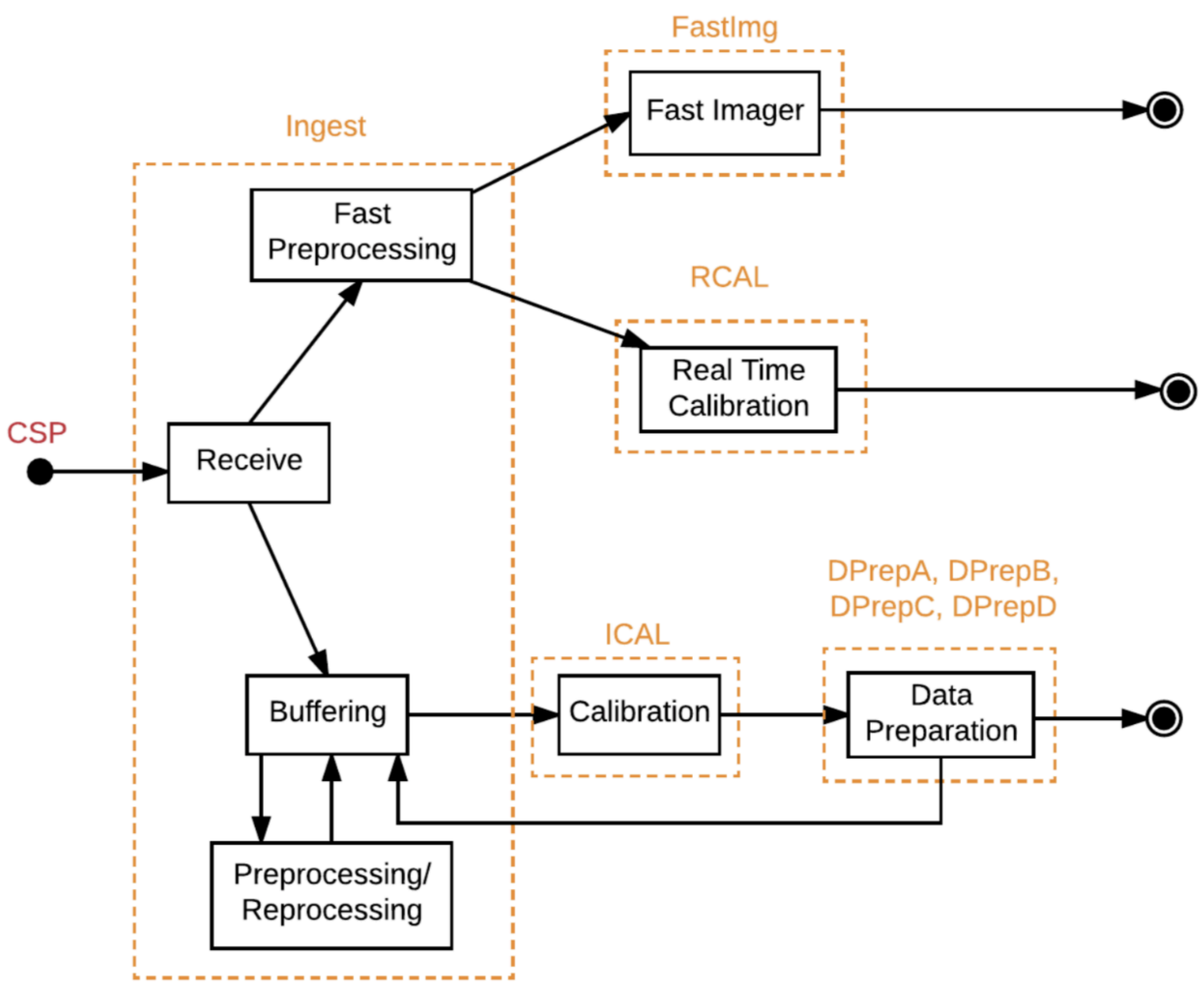

Figure 1: Overview of SDP pipelines [A07].

Minimum processing required to meet the needs of the HPSOs: **~26 PFlops**

This is subject to **strong** assumptions:

- Re-processing / post-processing is performed only **once** for each primary data product;
- That processing of SDP data products for all HPSOs is performed **at the same rate that they are ingested** into the SRCs

D3.2 Software Infrastructure

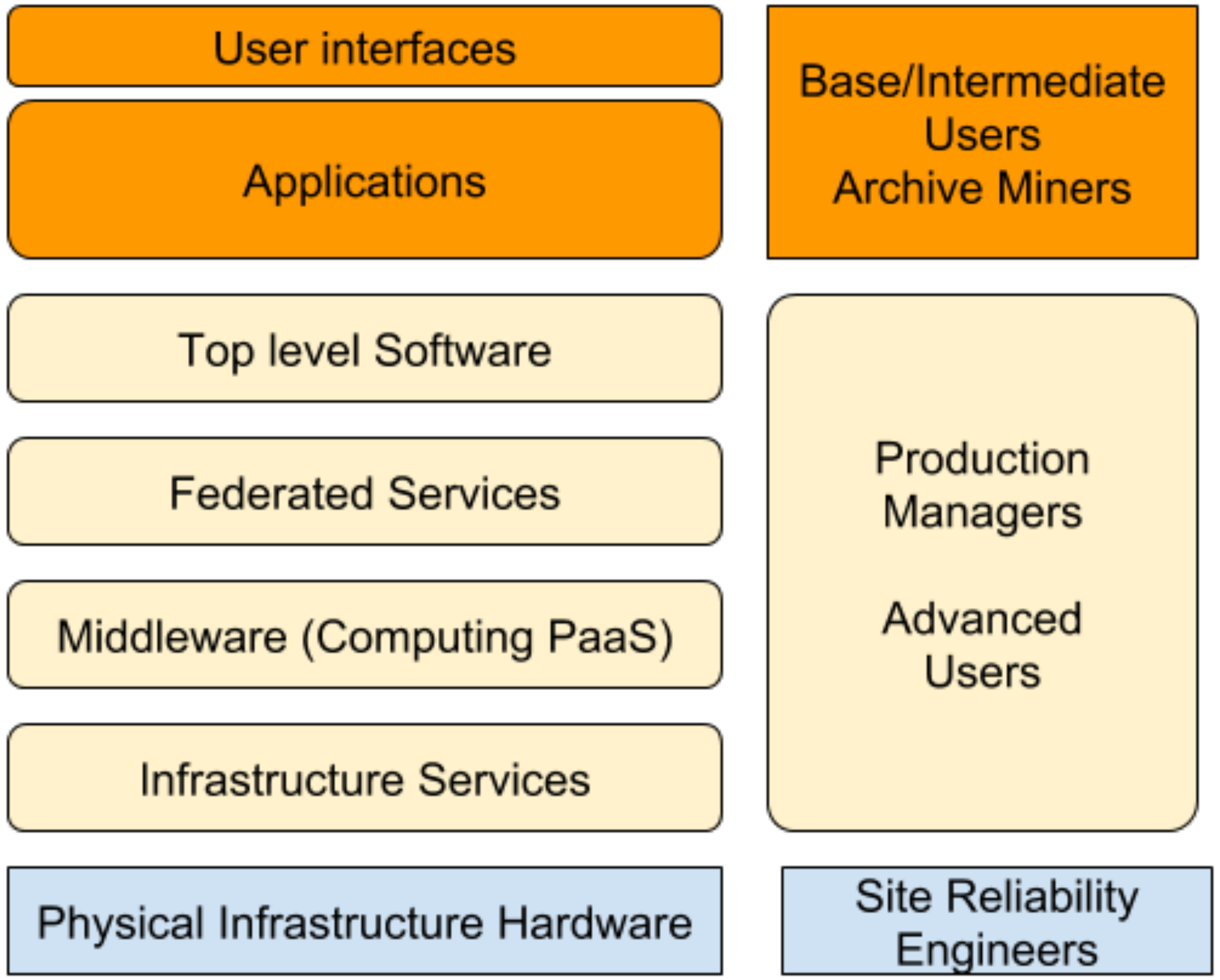

Figure 2.1: ESDC software stack and users' interaction.

- **Top-level Software** layer comprises all software components users interact with to their distributed applications
- **Federated Services** layer consists of all the components expected to be provided by each site to support SRC federation
- **Middleware (or Computing/PaaS)** layer is the set of orchestration and support services necessary to support distributed compute workflows within SRC infrastructure
- **Infrastructure Service** layer includes basic services for setting up and maintaining a computing site infrastructure

Identified four categories of users: basic, intermediate, advanced, and production managers, and made an assessment of their requirements

Some layers in more detail

- Applications (radio astronomy software)
  - SDP pipelines, precursor/pathfinder software, legacy software
  - Generic image analysis tools (source finders…)
  - Tools for particular science goals (RM synthesis, EoR power spectrum estimation…)
  - Visualisation tools
- Top-level software
  - Programming frameworks for running workloads on distributed hardware
  - Exposed to developers who want to write a new pipeline/application or modify an existing one
  - Technologies: MPI+X, Spark, Dask, …
  - Need ways to interact with the data: Jupyter notebooks, …
- Infrastructure services
  - Container technologies (Docker, Shifter, Singularity, …)

D3.3 Preliminary System Sizing

- Based on the computing and storage requirements identified in D3.1
- SRC sizing and costing follows same principles as the size and cost estimate for the SKA Science Data Processors
  - Facilitates implementation of the SDP computing models at the European SRC
  - Ensure consistency in projecting computing costs throughout the SKA project
- Assumes price date of 2025-02-01

Projection of costs

The approach to the projection of costs follows the model adopted by the SKA Science Data Processor (SDP), which has been guided by the following methods

1. Accelerator costs are projected based on a quasi Moore's Law base-lined on the prevailing GPU and CPU enterprise products at the reference date (2017-12-01);
2. Host CPUs are based on the current Intel Xeon processor family;
3. GPUs are based on the current Nvidia processor family;
4. Memory costs are projected based on an equivalent growth rate;
5. Network costs are based on a constant price per port and assume that the prevailing performance will increase driven by general market demands;
6. Storage costs are based on the prevailing Cost/GByte and projected based on industry-held growth-rates.

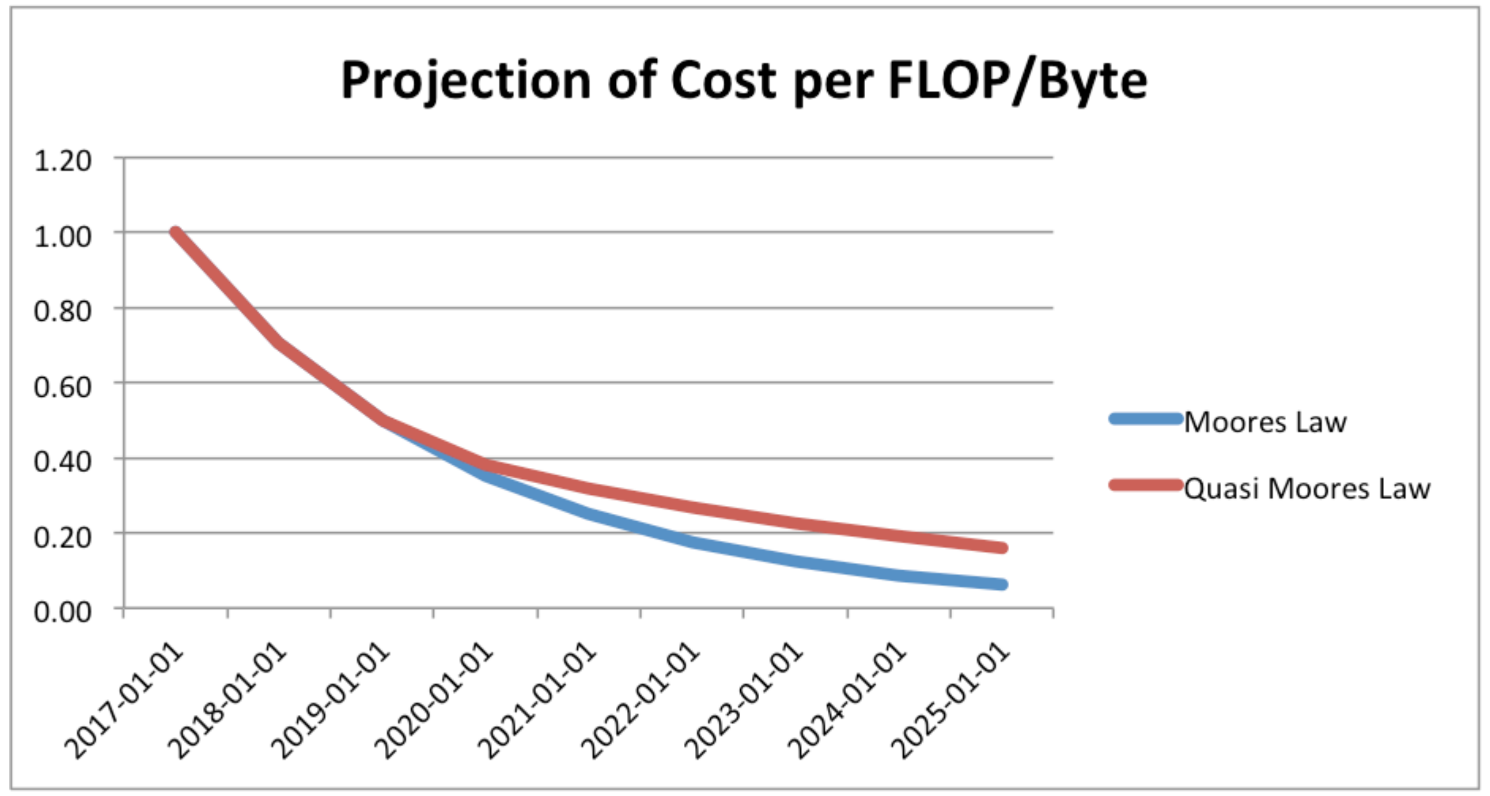

Figure 1: Quasi Moore's Law used for projection.

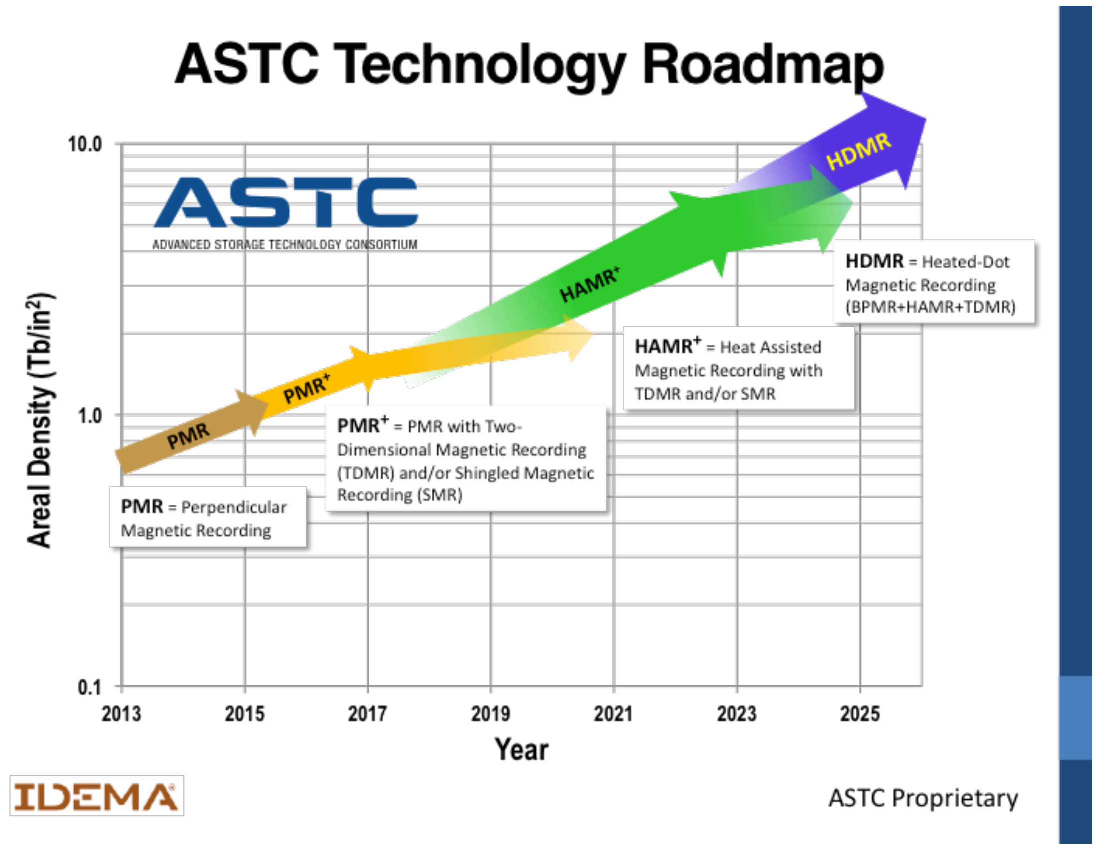

Figure 2: ASTC technology roadmap for storage density.

Compute Racks

- Compute racks form the basic unit of the SRC
  - Self-contained, independent collection of servers and infrastructure needed to keep the compute rack running
- Assume rack contains
  - 43 processing servers
  - 2 (intermediate) storage servers
  - 2 service servers
  - 1 remote service node
- Consider two styles of processing server: HPC and HTC

HPC Style Processing Server

Item	Quantity
Xeon Bronze Skylake	2
Accelerator	2
DRAM	64 – 1024 GBytes
25GbE	2
IB 100GbE	1
10GbE Cable	3
Power	~900 Watts

HTC Style Processing Server

Item	Quantity
XeonPlatinum Skylake	8
DRAM	64 – 1024 GBytes
25GbE	2
IB 100GbE	1
10GbE Cable	3
Power	~1200 Watts

Table 3: ESRC Processing server definitions.

Variation of costs with assumptions

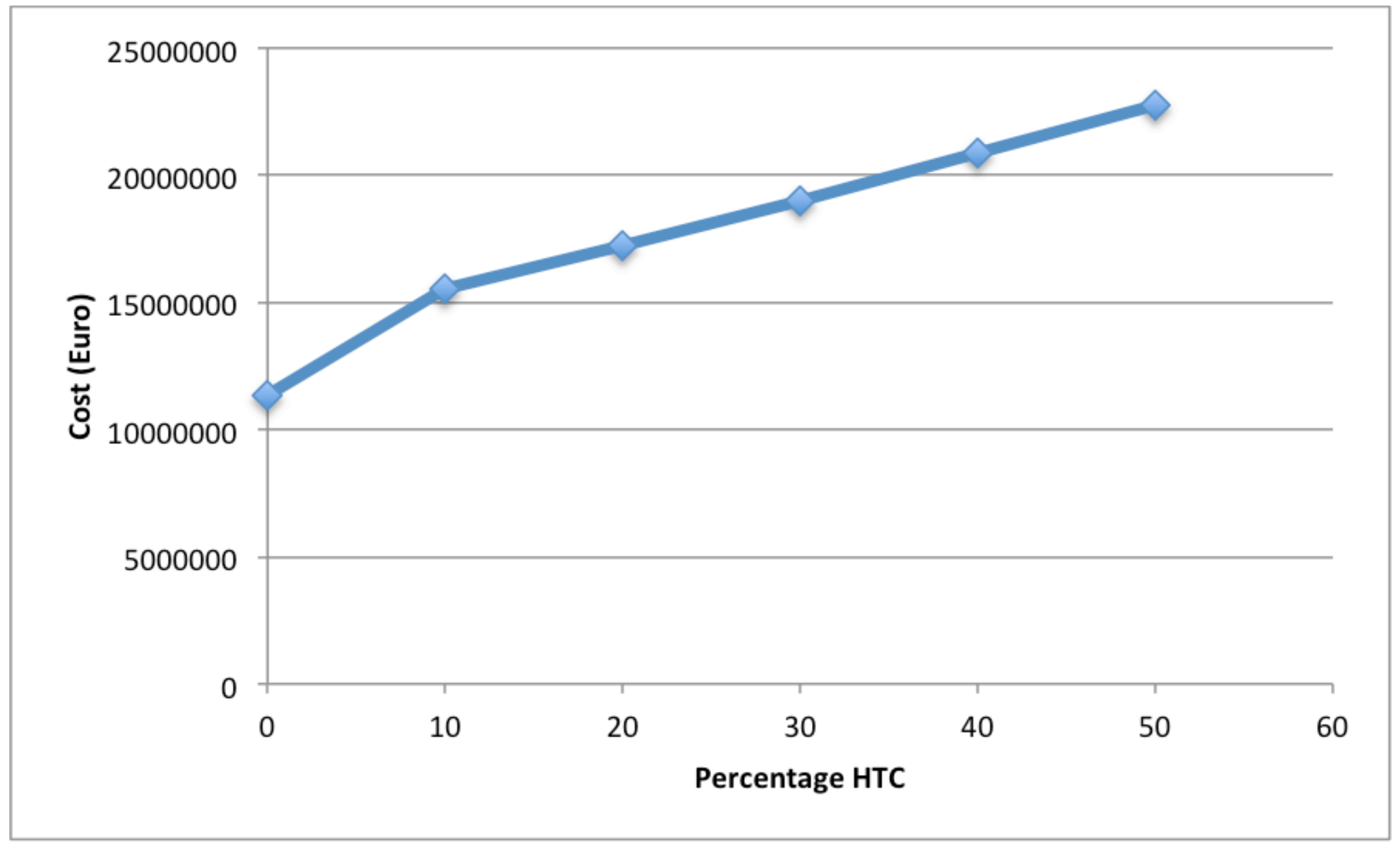

Figure 3: Variation in cost with percentage of total compute performed on HTC processing servers.

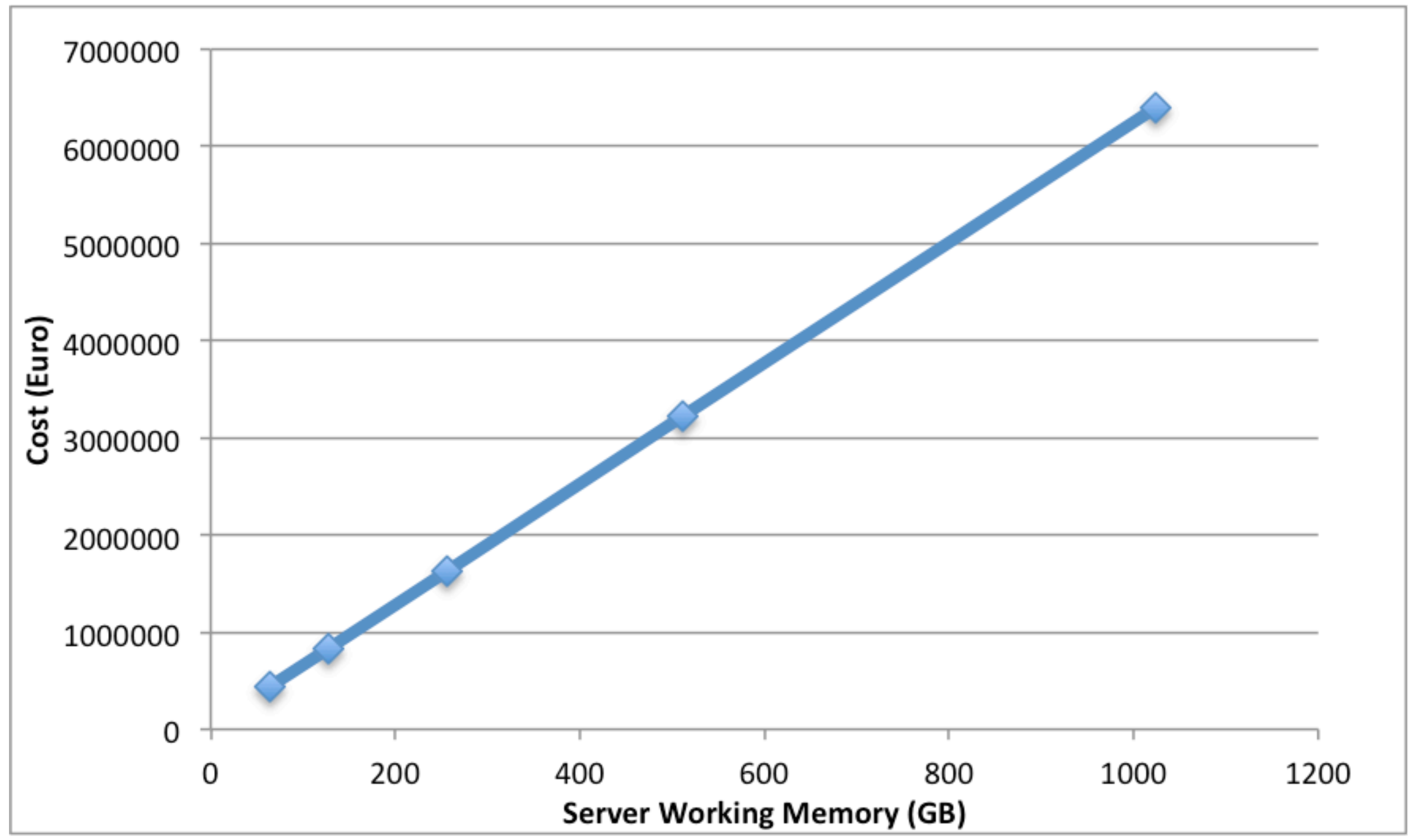

Figure 5: Cost as a function of Working Memory.

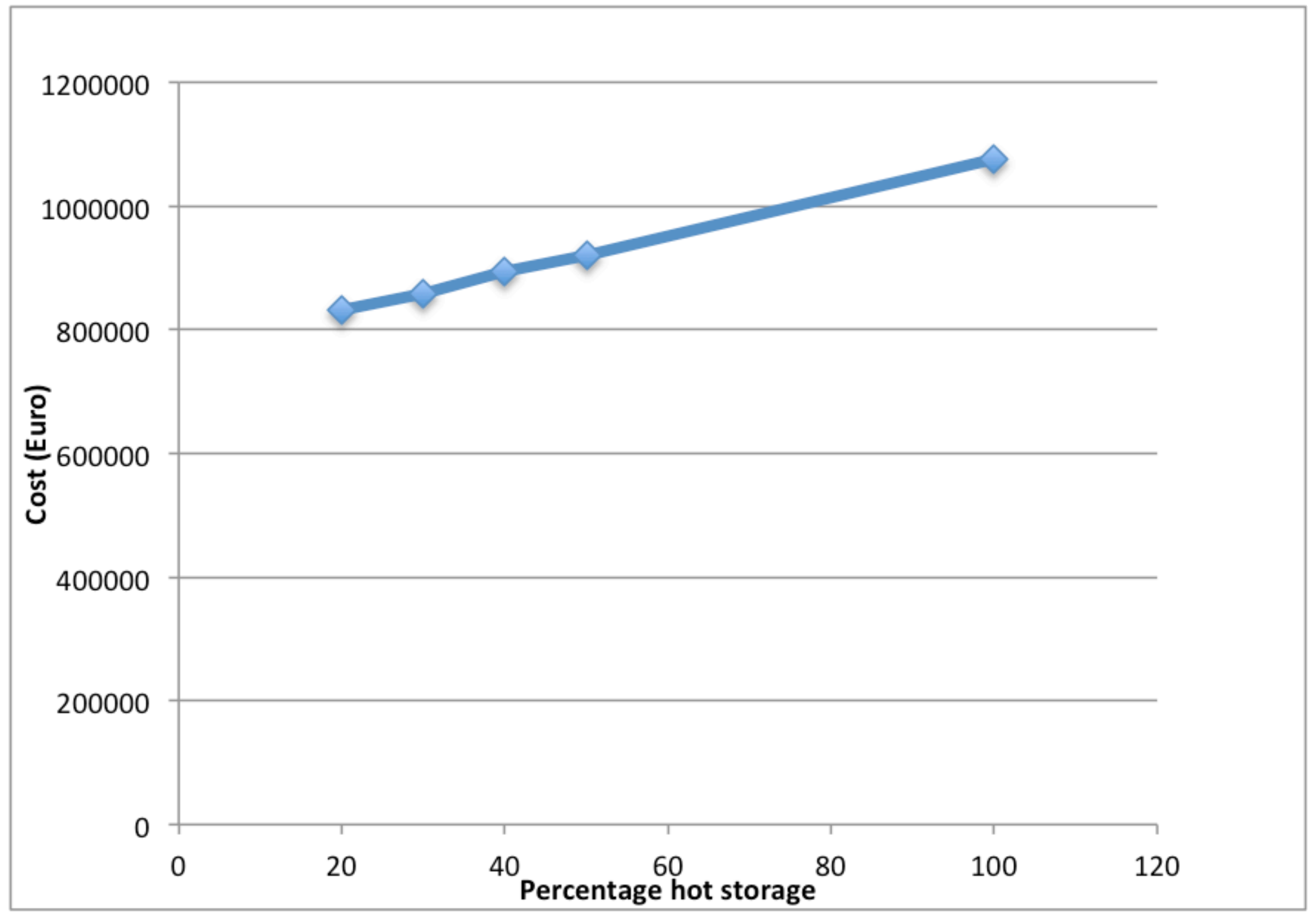

Figure 6: Cost as a function of hot storage in buffer.

Basic assumption:
Servers are all HPC

Basic assumption:
64 GB memory per server

Basic assumption:
25% storage in “hot buffer”

Cost Estimate

	Capital Expenditure for ESRC								
	Base Cost Estimate, excl.	Contingency drivers					Contingencies %	Cost Estimate, incl. Contingency	
	SRC	Tech Factor	Tech %	Cost Factor	Cost %	Sched Factor	RISK	SRC	(Sub) Totals
TOTALS	€27,987,500							€31,854,412	**€31,854,412**
Compute Hardware	€13,926,012						18%	€16,386,775	**€16,386,775**
Compute Rack									
Compute Rack Processing (excl. working memory)	€10,693,235	4	2%	6	1%	4	18%	€12,618,017	**€12,618,017**
Compute Rack working (pool) memory	€437,539	4	2%	6	1%	4	18%	€516,296	**€516,296**
Compute Rack Spares	€ 106,932.35	4	2%	4	1%	4	16%	€124,042	**€124,042**
Buffer Storage	€838,246	3	2%	4	1%	4	14%	€955,600	**€955,600**
Buffer Storage Spares	€ 8,382.46	3	2%	4	1%	4	14%	€9,556	**€9,556**
Infrastructure (Racks, etc)	€113,925	1	2%	6	1%	4	12%	€127,596	**€127,596**
Interconnect System									
Low Latency (inter-rrack) Core	€1,329,211	4	2%	6	1%	4	18%	€1,568,469	**€1,568,469**
Low Latency Network Spares	€ 132,921.08	4	2%	4	1%	4	16%	€154,188	**€154,188**
Management Network (incl in CI)									
Bulk Data Network	€210,932	4	2%	6	1%	4	18%	€248,899	**€248,899**
Bulk Data Network Spares	€ 21,093.16	4	2%	4	1%	4	16%	€24,468	**€24,468**
Preservation and Delivery Network	€30,542	4	2%	6	1%	4	18%	€36,039	**€36,039**
Preservation and Delivery Network Spares	€ 3,054.18	4	2%	6	1%	4	18%	€3,604	**€3,604**
Storage Hardware (per year)	€14,061,488						10%	€15,467,637	**€15,467,637**
Long Term Storage	€9,801,867	1	2%	4	1%	4	10%	€10,782,053	**€10,782,053**
Long term storage staging (HSM)	€4,259,621	1	2%	4	1%	4	10%	€4,685,584	**€4,685,584**

Table 1: Current cost estimate.