Metaheuristic layout design of a 2 billion euro science facility

P M Bentley and U Filges
1 European Spallation Source, Odarslösvägen 113, 225 92 Lund, Sweden
2 Forschungsstrasse 111, 5232 Villigen PSI, Schweiz
3 Author to whom any correspondence should be addressed.
E-mail: phil.m.bentley@gmail.com, philip.bentley@ess.se and uwe.filges@psi.ch

Abstract
A conceptual design process is described, whereby the layout of a modern, multi-MW spallation neutron source is optimised using metaheuristics. This was achieved by coupling several well-known analytical approximations of positive and negative performance metrics with a genetic algorithm (GA) and particle swarm optimisation (PSO). This will be contrasted with a parallel, manual study that was initiated immediately after. Overall, the metaheuristic method converged on superior results, and was generated with an order of magnitude reduction in total labour relative to the manual study. The baseline described in this report has become the real-world layout at the European Spallation Source (ESS), currently under construction in Sweden.

1. Introduction
Spallation neutron sources are large-scale science facilities for the study of condensed matter and fundamental physics. MW scale facilities, at the current forefront of the technology, fall into Flyvbjerg’s definition of megaprojects [1], and one such project under construction in Lund, Sweden is the European Spallation Source (ESS).

There are three major components to such a facility. The first component is a proton accelerator, which provides relativistic protons to illuminate the second component: a high-Z material target. Within the target volume, the exchange of gluons between quarks in the relativistic protons and target nuclei creates cascades of particle showers [2]. The resulting particle cascades liberate neutrons efficiently from the target material. The target assembly also contains hydrogen-rich, cryogenically-cooled neutron moderators [3] to inelastically shift the energy range of the neutrons into a scientifically useful window.

The third major component is the ‘business end’ of the facility: the suite of neutron scattering instruments [4]. The design of the layout of the instruments, and not the instruments themselves, is the subject of this article. The ESS aims to deploy 16 neutron scattering instruments in the first instance, with a foreseen suite of 22 instruments before 2030. These instruments are to be arranged in four sectors, named after the cardinal directions. The north and east sectors are intended for short instruments, less than 50 metres or so in length. The south and west sectors are intended for mostly long instruments.

The length of the instrument is important. The neutron creation process is timestamped, as is the detection event of a neutron at the instrument position [5]. It is the timing of the neutron detection that allows scientists to establish how fast the neutron was moving, which is directly related to the wavelength of the neutron and its kinetic energy [6]. Measurement of these properties can then be used to deduce wavelength- and energy-dependent phenomena using long-established experimental concepts [7, 8].

The timing of the neutron beams is therefore put under continuous computer control. In addition to the monitoring of the source pulse, precisely-phased rotating disks coated with neutron-absorbing materials—known as ‘choppers’—are used to control the neutron beam openings as a function of time, at multiple locations along the neutron beam-line. Their openings and rotation speed allow great flexibility in customising the instrument to the region of interest in parameter space. These choppers are typically a few tens of cm in radius, and up to one or two metres in total diameter at low rotation speed and the extreme limit of current practice.
In the north and east sectors, due to the size of the instrumentation being placed close to the source, it was always foreseen that some beam ports could not be deployed due to overcrowding. On the other hand, in the south and west sectors it was anticipated that every beam port could be used. In the north and east sectors, the beam ports are separated by 6°. In the south and west, the beam ports are separated by an alternating pattern of 5.3° and 6.7°. There are 32 beam port openings in total, with 8 ports per sector.

The task to be solved, therefore, is the identification of the optimum allocation of each instrument to a beam port, so that the suite as a whole maximises the performance. In decreasing congestion, complexity and interferences are the sectors of the west, north and east. This report will address the sectors in that order. The south sector is fairly trivial with only two firm instrument plans, and this sector was finalised by management without any study.

This kind of work is typically performed by negotiation between the instrument projects, with some kind of management committee as the deciding body that allocates the instruments to a particular beam port. In so doing, it is inevitable that the problem is simplified into a tractable form. This might be by ignoring complexities that otherwise might be an opportunity to improve the solutions, or making compromises that are otherwise not necessary. Another risk in this process is to the budget and schedule, caused by delays on decision making if attempts are made to inject technical details into the committee considerations, or time consuming debate continues between two or more parties over prime locations for critical hardware. Usually this results in option studies, evaluations, and reports. The risks are not negligible. To maintain control on costs, it is fairly well known that ‘implementation phases should be kept short and delays small’ as pointed out by Flyvbjerg [1], and he cites several sources with cost rates for blown schedules measured in M€ / month, similar to ESS. Complex strategic decisions of an architectural nature, meaning the functional vision of the system as a whole, should be resolved quickly using computational tools rather than slowly by large committees.

In light of the author’s previous work using metaheuristics at similar facilities [9–11], it was agreed that the approach should be attempted for this problem, since all instruments are unequal in their impact on neighbouring beam-lines. All established neutron scattering facilities have experience of dealing with some challenging interface problems—typically it is background effects transmitted as stray magnetic fields, noise from external radiation sources or electronic interference; encroachment of hardware into inconvenient spaces; or bottlenecks in access to nearby support infrastructure. At spallation sources, it is also the case that the energy spectrum at the source, especially the high energy components, varies significantly with angle relative to the proton beam, and therefore each beam port is not equivalent. The high energy part of the source spectrum is one source of instrument background that should be addressed early for MW spallation sources, since whilst it is possible to make some gains during operations [12] it is more effort- and cost-intensive than during the design phase. It will become clear in a moment that the stray magnetic field risks are already mitigated, so we can focus purely on the complexities of physical footprints and particle backgrounds. Capturing the full intricacy of potential interfaces between combinations of neighbouring instruments is theoretically possible, but it vastly expands the parameter space to be searched. In this light, it seems somewhat naïve to attempt a manual solution, which is why complexities of this scale have been simplified and/or ignored in the past.

Fortunately, soon after the decision to proceed with metaheuristics, a small team from a different department was indeed created and tasked expressly with the purpose of doing just that: to attempt to create a competing solution by hand. This presented an ideal and rare opportunity to compare human- and artificial-intelligence approaches on a megaproject. It will be shown that the simple AI methods proved to be superior.

2. Initial configurations

Figures 1, 2, and 3 show the working baseline, at the time that this work started, of the west, north and east sectors respectively (note that the instrument label ‘CAMEA’ has since been changed to ‘BIFROST’). It was immediately identified that the baseline was not viable due to mechanical interferences of the choppers with neighbouring beam-lines and between instrument infrastructure and the buildings.

3. Description of the optimisation problem

The optimum beam allocation, within each sector, is the selection of an optimum permutation of 12 beam port allocations, where a beam port can be allocated a single instance of a unique instrument, or be allocated a status of empty. The number of permutations of 12 beam ports is given by the well known equation:

$$P_k = \frac{n!}{(n - k)!}$$  \hspace{1cm} (1)

which is the number of different, ordered subsets of size $k$ taken from $n$ items.
With 32 beam ports available to 22 instruments, it might seem that there are $7 \times 10^{28}$ possible solutions, however the geography of the ESS site already partitioned the instruments into one long group on the west sector, one medium group on the south sector, and two short groups on the north and east sectors. In the long group, the instrument end stations are sufficiently far away that stray magnetic fields at that position are not likely to be problematic. For the two short groups, the separation of strong magnetic fields from instruments that are sensitive to stray fields had already been performed, so we can ignore this complexity for the remainder of the study.
After this geographical and magnetic partitioning, a more precise description of the permutations of the discrete parameter space is given in table 1.

These could be solved by a brute-force exploration in an acceptable time frame, but there are other continuous parameters to be modified. One example is physical geometry: there is some flexibility available on the position of the first chopper. The physical location of the first chopper defines the bandwidth of the instrument, and moving it away from the source amounts to a sacrifice of bandwidth. As will be clarified shortly, this may or may not be a problem, depending on the instrument design. In the solution reported here, the positions of the first choppers are continuously variable between 0-1 m, and the optimum would be expected to stagger the first chopper placements in some way between the neighbours. The parameter space is therefore much larger than the simple permutations summarised in table 1, and the optimum beam port locations is correlated with the staggered deployment of the choppers.

One could tackle this problem by treating it as a mixed integer linear programming problem (MILP) using commercial [13, 14] or open source [15, 16] tools. The reason that this option was not taken was that the final scope of the problem, as presented here, was not clear at the outset, and non-linearity was foreseen in the problem. The metaheuristics already had working interfaces to more advanced modelling tools, to vary the beam angle, supermirror value and radius of curvature if needed, as was necessary in the past on a similar project [11].

![Figure 3. Working ESS baseline layout of the east sector, before the optimisation work. Red = mechanical interference. The radius measurements are in metres.](image)

Table 1. Size of the discrete parameter space for each sector, ignoring continuous parameters described later in the text, where \( k \) is the number of instruments in each sector.

| Sector | \( k \) | \( \Omega P_k \) (Rounded) |
|--------|-------|--------------------------|
| North  | 5     | 95 000                   |
| East   | 6     | 665 000                  |
| South  | 2     | 130                      |
| West   | 8     | 20 000 000               |
We therefore turn to more elegant computational methods, based on artificial intelligence emerging from the collective interactions of simple agents. It will be an interesting future project to compare results of a simpler MILP approach with those that were obtained from the metaheuristics.

3.1. Implementation

Metaheuristics are general purpose optimisation algorithms with the aim of improving solutions towards an optimum, but without guarantees that a global optimum is found. These can be compared to traditional analytic methods, which attempt to encapsulate the salient features of the problem in a mathematical model, and are hard-coded to solve a particular problem; and iterative optimisation strategies such as the Newton-Raphson method, which do not tolerate noisy parameter spaces with multiple local optima.

Metaheuristics form part of the basic artificial intelligence and data science toolkit, and are now established methods with decades of widespread use. These tools are especially useful in large, complex problems where traditional approaches fail, because there are multiple solutions; the parameter space is not differentiable, or is noisy; or where it is difficult to define a complete mathematical model of the positive performance metrics.

The two algorithms used in this study draw on the collective intelligence that emerges from the behaviour of interacting agents. This is evolution-based, in the form of a Genetic algorithm (GA) [17]; and social insect-based, in Particle Swarm Optimisation (PSO) [18]. In these methods, a population of solution candidates are evaluated and interact with each other to improve the population as a whole.

Algorithms inspired by complex biological and physical processes are numerous, and there are many that fall into the metaheuristic category. Simulated annealing (SA) [19], ant colony optimisation (ACO) [20], and differential evolution (DE) [21] are popular methods that are grouped under adjacent categories as GA and PSO. ACO is designed to solve optimal graph traversal. DE and SA are valid alternatives to GA and PSO. The advantages of GA and PSO are both subjective and objective: in the former, the author finds GA and PSO highly intuitive and easy to customise; and in the latter a fairly broad study by E. Farhi [22] benchmarking multiple methods available under a semi-commercial platform found that PSO was the most efficient in noisy parameter spaces that occur when modelling neutron beams. As mentioned in the previous section, the tolerance of PSO and GA to noisy, non-linear parameter spaces were foreseen at the outset as a likely essential feature.

Which metaheuristic method is best suited to a particular problem depends on the problem space and the performance surface geometry. Subjectively, the author has found that PSO tends to find the deeper areas of sharper performance maxima better than GA, whereas GA escapes from local optima more readily than PSO due to the GA’s mutation operator causing potentially larger changes in parameter values than can be generated in a stable fashion in PSO. GA appears to be better suited to optimising some classes of sequential problems compared to PSO. It only costs a few additional minutes of CPU time to deploy them both under a common framework, as was done in this case.

In GAs, illustrated in figure 4, the population improvement mechanism is the breeding of the best solutions to generate successive generations, so as to converge towards the desired behaviour. A population of 50 agents was created with 32-bit, normalised Gray-coded chromosomes; rank selection was used at each generation; the mutation rate was set to 0.07; uniform crossover was set to 0.8, meaning that 20% of the rank selected population is cloned into the next generation, and in 80% of the children a bit-by-bit lottery is used to select equally the rank selected population.

In PSO, illustrated in figure 5, this involves the solutions moving through the high-dimensionality parameter space, and accelerating towards the location of the most recent global optimum \( P_G \) that is communicated through the whole swarm, and towards its own optimum location \( P_i \) that remains secret to each agent. The agents are moved in unit time steps according to the following equations:

\[
x_{i+1} = x_i + v_i
\]

\[
v_{i+1} = Dv_i + C_L R_L(x_i - P_i) + C_G R_G(x_i - P_G)
\]

\[D = 0.95\] is a damping factor that slows the particles down, \( C_L = C_G = 1 \) are local and group weighting factors (and thus can be ignored in this case) and \( R_L \) and \( R_G \) are random weighting factors in the range \( 0 \leq R \leq 1 \) that are updated at each iteration. The resulting movement of solutions in PSO strongly resembles the oscillatory search patterns of foraging wasps, such as *vespula vulgaris*, that are typically observed at picnics close to a fragrant food source such as fermented drinks, meat or cakes.

One can tune the parameter values for GA and PSO to improve the performance for each specific optimisation problem type. However, this is a ‘single-shot’ study, unlike in finance or business strategy we do not need to repeat the calculation on a regular basis. The behaviour of the algorithm is somewhat insensitive to these
parameter values; and these remain largely unchanged from previous work [23]. The primary motive is to achieve a superior optimum in a significantly shorter time compared to a human being.

It should be clear that the smoothness or differentiability of the parameter space therefore doesn’t matter, and difficulties to define precisely a positive performance metric is replaced by more generic flags and the easier definition of negative performance penalties. For example, it can be tricky sometimes to write a precise list of Christmas gifts that one actually needs or wants whilst navigating the social minefield of potential costs. However, it is trivial to write a vague positive list (quirky sci-fi/fantasy board games, festive foods), along with a blacklist (no mystical trinkets or popular culture), thus enabling the agents to find viable solutions on their own.

In the present work, both GA and PSO methods were deployed but the superior solution was found by PSO. It is the same code base as was used on similar problems involving neutron instrument design [9–11, 23]—the only modifications were to add a 2D visualisation via Eukleides [24]; a 3D visualisation with POV-Ray [25], and to take advantage of extra CPU cores by multi-threading.

Figure 4. Flow diagram of a simple genetic algorithm.
The code terminates after 200 iterations, which was found to be optimal in previous work for both algorithms [23]. On an Intel i7-6700HQ CPU running at 2.60 GHz, PSO converges after 47 seconds, and GA after 124 seconds. The difference in timing is mainly down to the fact that the PSO converges efficiently with a population size of 30, whilst GA requires a few more evaluations with a population size of 50. Both of these timings are sufficiently short for the needs of the project.

3.2. Modelling the problem

The relative performance of each individual is quantified by a non-linear equation that is the sum of many penalty scores.

Management agreed on the following requirement prioritisation to guide the figure of merit:

(i) To provide the best outcome in terms of performance and operability, for instruments 1–15 and the two options for instrument 16

(ii) To provide the best outcome for ‘guessed’ instruments 17–22; and

(iii) To minimise blocking of unallocated beam-ports, such that 32 instruments (or more) could be installed.
Table 2. Fast neutron flux viewed through a T0 chopper (0°) and from neighbouring beam lines (5–15°), neglecting thermalisation processes (i.e. the total neutron fluxes are higher than stated here).

| Separation (°) | Flux (n/m²/s) | Relative BG Suppression |
|---------------|--------------|-------------------------|
| 0            | 1.2 × 10⁸    | —                       |
| 5            | 4.0 × 10⁶    | 1                       |
| 10           | 7.8 × 10⁵    | 1/5                     |
| 15           | 2.2 × 10⁵    | 1/32                    |

The prioritisation cascade is coded on a logarithmic scale, meaning that scores associated with item (i) are weighted × 10 those for item (ii), which are weighted × 10 those of item (iii), and so on for additional prioritisation. This logarithmic scale ensures consistent behaviour across large volumes of parameter space.

The performance priority (i) is subdivided into a number of scoring functions as follows:

- **Spallation Background** whereby the algorithm is encouraged to move background-sensitive instruments away from the A2T area in the North and East sectors, and select beam ports in directions that have reduced fast neutron intensity.
- ‘Noisy Neighbours’ where background-sensitive instruments are kept away from those likely to be sources of background (e.g. straight instruments, imaging etc.). This is guided by our own study of the SNS [26], knowledge from PSI, Switzerland, and private communication from G. Mührer, ESS.
- ‘Beam cross talk’ where the distance between two chopper pits, or chopper pits and guides, is maximised to reduce cross talk and mitigate both safety and background risks/costs for the suite.

The beam cross talk is guided by MCNP results which are summarised in table 2. The cross talk location is close to the target monolith as a function of beam port angular separation, with a small amount of lateral shielding as space allows. The tally location is at 16 metres. The final column is the relative cross talk suppression relative to the minimum 5° beam separation. The bunker is fully shielded in these models, maximising albedo transport at high energy. This is a guide for the instrument cross-talk levels in a typical bad-case scenario, and it does not correspond exactly to the geometry of any single instrument concept.

The results in table 2 show that the cross talk strength has a significant reduction with increasing separation, as expected. It is also worth noting that, for a pair of instruments separated by 5°, the cross talk from the neighbour is roughly 1/30 relative to the flux transmitted through the T0 chopper. Depending on the instrument geometry, this may or may not be significant.

No assumptions were made about the conceptual design of the instrument caves, except to reserve a thick boundary around the detector space that is commensurate with estimated beam intensities and expected shielding specifications based on concepts developed as part of larger collaborations [27]. It is assumed that any access points would be resolved during detailed design, based on the project interfaces generated by this baseline. The logic at work is that it makes no sense to allow a non-essential feature appearing in a CAD drawing to be escalated to a priority (i) issue because it causes a mechanical interference, and thus sacrifice other benefits that would be found by ignoring such details at present.

The next step is to apply boundary conditions. This can be done parameter by parameter, but more conveniently by defining a simple level (0) prioritisation (surpassing the above list) that penalises very heavily any solutions that are non-viable. It is often quicker to write a simple Boolean function that checks whether a solution is valid than to define all valid solutions.

The final equation for performance of n = 8 instruments allocated to p = 8 ports is given as an objective function that must be minimised (a penalty function), as follows:

\[
p = w_0 \sum_{i=1}^{n} b(i) + w_0 \left( \sum_{i=1}^{n} k(i) + \sum_{i=1}^{n} \sum_{j=1}^{p} mn(i, j) \right) - w_1 \times (tf1(i) + tf2(i)) + w_1 \times \sum_{i=1}^{n} s_ir(a_i) + w_1 \times \sum_{i=1}^{n} t(i)
+ w_1 \times \left( \sum_{i=1}^{n} \sum_{j=1}^{p} s_{i,j}(n)ab(a_i - a_j) + \sum_{i=1}^{n} v_i c_j \right) + w_3 \times \left( \sum_{i=1}^{n} y(i) - tf3(i) - tf4(i) \right)
\]

where the summations are only evaluated for instruments and beam ports that are assigned, and the vector \(a_i\) gives the beam port assignments of each \(i\)th instrument. \(w_0 = 100000, w_0 = 10000, w_1 = 1000,\) and \(w_3 = 10.\) \(w_2 (=100)\) was coded but not used, as it was found that in the summations a larger separation was needed. \(b(i)\) is a function returning 1 if the boundary conditions are not met for the \(i\)th instrument; \(k(i)\) is a function returning the count of mechanical interferences for the \(i\)th instrument with surrounding buildings; \(mn(i, j)\) is a function.
returning the sum of mechanical interferences between instrument $i$ and the instrument on port $j$; $tf1()$ is a function that counts the number of ‘tranche 1’, high priority instruments have been allocated a beam port; $tf2()$ is a function that counts the number of ‘tranche 2’, lower priority instruments allocated to a beam port; $s$ is a vector of relative instrument sensitivities to background noise based on the instrument class, measurement timing and detector geometry; $r()$ is a function returning the relative spallation source fast neutron background strength, which varies with the angle relative to the proton beam, and hence the beam port allocation; $lt(i)$ is a function only used for the short sectors, which quantifies the minimum thickness of shielding available for the $i$th instrument; $n(j)$ is a function returning the relative noise level of the instrument located on the $j$th beam port, based on its geometry, and $abs(a_j - a_i)$ is a positive quantity proportional to the spatial separation of the two instruments; $v_i$ is a vector of relative penalties for each of the $i$ instruments per unit of translation of the first chopper, as given in table 3; $c_i$ is a vector of first chopper positions, i.e. each element is a continuous variable for each instrument; $y(i)$ counts the number of mechanical interferences between the $i$th instrument and any empty beam ports (preventing future use); $tf3()$ returns the number of ‘tranche 3’ lower priority instruments allocated to a beam port; and $tf4()$ returns the number of ‘tranche 4’ empty beam ports that are allocated to a beam port.

The code attempts to allocate more instruments than there are beam ports available, and the prioritisation above resolves the over-subscription in an optimum way.

### 4. Results

#### 4.1. West sector

The initial runs showed quite clearly that, with the available data [28], the HEIMDAL instrument would need to occupy the space of two instruments in the long sector. This is incompatible with the ESS objectives, and so management decided to constrain the instrument to one instrument footprint, and reduce the size of the second cold chopper ‘CT2’ from 1.36 m down to 0.8 m.

It was not possible to find a solution that required no changes to the positions of the instrument choppers. All of the solutions required one or more instruments to sacrifice bandwidth, so the effort then switched to a weighted search as described in the previous section. The ESS management gave feedback as to what the impact would be on the science case for each instrument with a chopper close to the target, and what the priorities were. These are summarised in table 3. Here we can see that the penalty for shifting BEER’s bandwidth chopper, for example, would be $3 \times$ greater than the penalty for shifting MIRACLES’ chopper. Invariably, this results in a prioritisation queue for sacrificing bandwidth.

The solution found with these weights was actually superior to those produced with less guidance on which instrument to penalise. This is a reflection on the size of the parameter space and the complexity of the problem.

At this stage, management returned with additional requirements, and the strength of this kind of approach becomes clear. Adding a few extra lines of code and re-running the optimisation takes only a few hours, compared to manual studies which would have to be repeated over several days, or perhaps weeks. The new requirements were to place NMX on the first port in this sector, and to place BEER close to essential infrastructure, namely in any of the next three ports. With these requirements implemented, the suite was solved as shown in figure 6, where we can see that NMX and HEIMDAL remain in the positions chosen by human reasoning, but those beam-lines in between have been rearranged. An observant reader will note that the instrument designs were not frozen during this study, and some changes can be seen in the chopper sequences when comparing with the baseline in figure 1.

#### 4.2. North sector

In the north sector, the baseline issues were interference between SLEIPNIR and WANSE, FREIA and a potential future beam port, and an arbitrary allocation considering background issues.

### Table 3. Weighting factors, $v_i$, for penalties of moving choppers.

| Instrument | Impact |
|------------|--------|
| MIRACLES   | Low (1) |
| MAGIC      | Medium (2) |
| BIFROST    | Medium (2) |
| HEIMDAL    | High (3) |
| BEER       | High (3) |
| Default    | 0      |


The optimal solution that was found is shown in figure 7. Under management instructions, the mechanical interferences with NNBAR and WANSE were relaxed. Both of those instruments were seen as potential future instruments, and at the time of this work management stated a strategy of building NNBAR using external funds,
for a specific duration, and in the long term replacing NNBAR with WANSE. This is reflected in the layout as shown in the figure. It should also be clear that FREIA’s chopper diameter is larger in two locations than other beam-lines, creating an unavoidable interference that always blocks one beam extraction. In this case, the third empty beam-line is obscured. Lastly, we can see the effect of minimising cross-talk and background, by moving the instruments as far as possible away from the proton beam axis and spacing the instruments apart. The end result is a removal of interferences on SLEIPNIR and WANSE, a swap of LOKI and FREIA, and spacing the instruments as far apart as possible.

4.3. East sector

In the east sector, the two issues with the baseline were an interference problem between VOR, the building, and its neighbouring instrument, ESTIA; and an arbitrary placement from a background point of view.

An optimised solution is shown in figure 8. There are no longer any mechanical interferences, and the placement of the instruments provides a greater avoidance of background.

5. Discussion

The layout presented here became the working baseline for the ESS [29]. There were a number of minor change requests that were approved after management reviews [30–35], but they do not affect the logical outcome of the present work and were made in light of more details being developed in the instrument concept.

In terms of workload, it took perhaps 3 weeks of intensive work to create the core functionality of the software. Occasional interaction with a mechanical engineer provided the standard footprints of classes of components, along with the detector and shielding envelopes for all the instruments. This support was so light and infrequent that it was not considered important to quantify it on the project time-sheets.

The manual study relied on the same mechanical engineer, allowing a comparison to be made. The manual study required many, multi-hour meetings and was described as almost full time work of several person-months to integrate all the CAD drawings of all the instruments. The meetings were intensive discussions of possible permutations and compromises, between which was a tedious effort of moving items around and trying to resolve interferences by hand, generating and circulating multiple option sketches for feedback. This cycle of work was repeated week after week for months.
The final result of the manual study is documented [36]. There remain mechanical interferences, complex and unresolved trade-offs, and an overwhelming list of options and details that have not been concluded. Moreover, each clarification and modification resulting from meetings with management required a re-working to produce similar deliverables under more complex conditions. The feeling of being trapped in an endless cycle of repeating tedious lists of actions had a significant, negative impact on the morale and motivation of the engineering support, ultimately leading to the loss of a highly skilled engineer. It also acts as a sink of labour that would have been more usefully spent on other issues.

In contrast, the management change requests are much faster under the AI approach. Implementing new boundary conditions or adjustments to metrics requires modification of a few lines of C++ in a single file. Afterwards, the whole design process can be re-run in minutes with the new conditions. The engineering support was simply a 5 minute coffee meeting to ensure viability of the solutions. This meant that the turnaround time was so short that the follow-up meeting schedule was defined by social factors rather than labour estimates. On close out, a single CAD model is built to update the baseline to the solution defined in this report. Whilst the AI method can be (and was) re-run several times with a different random number seed, and thus also generate several alternative options, the evaluation metric is obtained with each option and this makes it easy to pre-select and discriminate between options in an objective and quantitative way. Hence, the stakeholders are able to review the proposed vision as a whole, and not find themselves drawn into lengthy negotiations about trade-offs, i.e. how the concept is being produced.

The final design eliminated all mechanical interferences, as well as sacrificing a small amount of bandwidth on an instrument that is the least sensitive to bandwidth reductions. At the same time, it separated background sensitive instruments from instruments that are expected to be sources of cross-talk, whilst also placing them as close as possible to optimum beam port angles with respect to the proton beam for source background contributions.

A final point of note is that, whilst this study intended to have a reduction on the ESS instrument backgrounds through intelligent deployment of the infrastructure, aggressive cost saving steps in other parts of the project may initially overwhelm the background reductions [37]. The stated management strategy is to identify and correct these during operations, so the gains may not be apparent until this remedial phase is fully complete. It is the opinion of the authors that this does not mean that the approach is invalid. On the contrary, any future projects would do well to think along similar lines, since the surrounding infrastructure may well differ; and the other benefits to this method are substantial enough to warrant recruitment of the required expertise to do likewise.

### 6. Conclusions

An agent-based metaheuristic method has been used to create the baseline layout of a ~2 billion euro megaproject, namely a leading MW spallation neutron source. It has been demonstrated that this is a more efficient and flexible approach compared to doing so by hand. The AI method is orders of magnitude quicker, less labour intensive, and more flexible to management change requests with a faster turnaround. Such approaches automate, and largely eliminate, the repetitive cycles of designing, presenting and reviewing in lengthy committee meetings, allowing the stakeholders to focus on the strategic vision rather than becoming trapped in the details of the optimisation process itself. This carries with it an improvement in staff morale, which should not be underestimated on large projects. The priority is not just the initial project delivery, but effective and long-term provision of highly specialised hardware of international importance, for which the retention of motivated and fully engaged experts is critical.

Complexities in megaprojects should rely from the outset upon numerical optimisation methods, either with metaheuristics or MILP approaches, and only use manual methods to validate. In such work, which of MILP or metaheuristics are the optimal strategy in the absence of Monte-Carlo simulations of beams depends on the parameter space and the evaluation methods used. It is an interesting and open question worthy of future study.

### Acknowledgments

The author is especially grateful to B Davidge and Z Lazic (ESS, Sweden) for engineering input and many interesting and constructive discussions.
ORCID iDs

P M Bentley @ https://orcid.org/0000-0003-1235-1011

References

[1] Flyvbjerg B 2014 What you should know about megaprojects, and why: an overview Project Management Journal 45 6–19
[2] Boudard A, Cugnon J, David J-C, Leray S and Mancusi D 2013 New potentialities of the liège intranuclear cascade model for reactions induced by nucleons and light charged particles Phys. Rev. C 87 014606
[3] Zanini L, Rutkov K, Klinkhy E, Mezei F, Schönfeldt T and Takihayev A 2018 The neutron moderators for the european spallation source Journal of Physics: Conf. Series 1021 012066
[4] Andersen K H et al 2020 The instrument suite of the european spallation source Nuc. Inst. Meth. Phys. Res. A 957 163402
[5] Mukai A H C et al 2018 Architecture of the data aggregation and streaming system for the european spallation source neutron instrument suite J. Instrum. 13 T10001
[6] De Broglie L 1925 Recherches sur la théorie des quanta Ann. Phys., Paris 10 22–128
[7] von Halban H and Preiswerk F 1936 Preuve experimentale de la diffraction des neutrons Comptes rendus hebdomadaires des séances de l’Académie des sciences 203 73–5
[8] Brockhouse B N and Stewart A T 1955 Scattering of neutrons by phonons in an aluminum single crystal Phys. Rev. 100 756
[9] Bentley P M, Pappas C, Habicht K and Lelièvre-Berna E 2006 Evolutionary programming for neutron instrument optimisation Physica B 385 1349
[10] Holzel R, Fouquet P and Bentley P M 2006 Instrument design and optimisation using genetic algorithms Rev. Sci. Inst. 77 105107
[11] Bentley P M, Boehm M, Sutton I, Dewhurst C D and Andersen K H 2011 Global optimisation of an entire neutron guide hall J. Appl. Cryst. 44 483
[12] Cherkaishyna N et al 2014 Overcoming high energy backgrounds at pulsed spallation sources Proc. of the XXI Meeting of the International Collaboration on Advanced Neutron Sources (ICANS-XXI) (https://doi.org/10.11484/jaeca-conf-2015-002)
[13] Bixby RE 1998 CPLEX (www.ibm.com/analytics/cplex-optimizer)
[14] Gu Z, Rothberg E and Bixby R 2008 Gurobi (www.gurobi.com)
[15] Mitchell S A, O’Sullivan M J and Dunning I 2011 PuLP: A Linear Programming Toolkit for Python Online - - (http://www.optimization-online.org/DB_HTML/2011/09/3178.html)
[16] Schumacher D Mixed integer linear programming inr (https://github.com/dirkshumacher/ompr)
[17] Holland J T 1975 & 1992 Adaptation in Natural and Artificial Systems (Cambridge, MA, USA: The MIT Press)
[18] Kennedy J and Eberhart R 1995 Particle swarm optimisation Proc. IEEE Int. Conf. Neural Netw. 4 1942–8
[19] Kirkpatrick S, Gelatt C D Jr and Vecchi M P 1983 Optimization by simulated annealing Science 220 671–80
[20] Colomi N, Dorigo M and Maniezzo V 1991 Distributed optimization by ant colonies Actes de la Première conférence Européenne sur la Vie Artificielle (ÉCAL’91, Paris, France) (Amsterdam: Elsevier) pp 134–42
[21] Storm R and Price K 1997 Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces J. Global Optim. 11 341–59
[22] Farhi E, Debbay Y and Willendrup P 2014 ifft: a new data analysis framework, applications for data reduction and optimization of neutron scattering instrument simulations with mcstas J. Neutron Res. 17 5–18
[23] Bentley P M and Andersen K H 2009 Optimisation of focusing neutron devices using artificial intelligence techniques J. Appl. Cryst. 42 217
[24] Obrecht C 2004 Eukleides (http://www.eukleides.org/)
[25] Persistence of vision (tm) raytracer. (http://www.povray.org), 2015
[26] Di Julio D D et al 2016 Characterization of the radiation background at the spallation neutron source Journal of Physics: Conf. Series 746 012033
[27] Zenderl C, Santoro V, Ansell S, Cherkaishyna N, Martin Rodriguez D, Di Julio D and Bentley P M 2015 ESS neutron optics and shielding handbook Technical Report, European Spallation Source ESS-0039408
[28] Christensen M, Lefnann K, Lindahl Holm S, Bertelsen M, Shefer J and Norby P 2014 ESS instrument construction proposal HEIMDAL Technical Report, European Spallation Source
[29] Bentley P M 2019 Baseline for ESS beamport allocation Technical Report, ESS-0047786
[30] Andersen K 2016 Change of SKADI beamport from E8 to E5 Technical Report, ESS-0055782
[31] Andersen K 2016 Response to SKADI baseline change request Technical Report, ESS-0058475
[32] Glavic A 2016 ESS beamport allocation baseline change request for ESTIA Technical Report, ESS-0052476
[33] Andersen K 2016 Response to ESTIA baseline change request Technical Report, ESS-0056731
[34] Andersen K 2016 Response to DREAM beamport allocation baseline change request Technical Report, ESS-0081039
[35] Andersen K 2016 Change of SKADI beamport from E5 to E3 Technical Report, ESS-0080327
[36] Strobl M 2015 Beam port deployment at ESS—a design study Technical Report, ESS-0048746
[37] Bentley P M 2020 Potential design issues for ESS initial operations phase in the sphere of optics and shielding in preparation