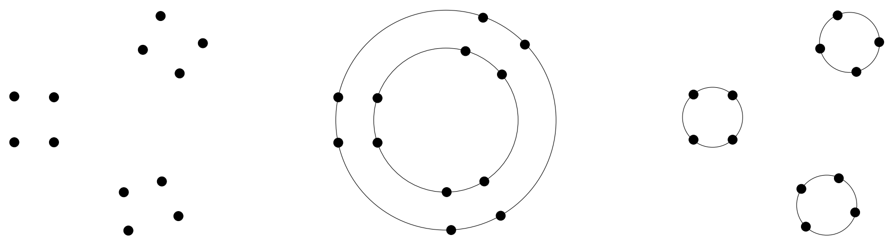

Figure 1: What rings should we see in (a)? Are there two large concentric rings as indicated in (b)? Perhaps there are three small rings of equal radii as indicated in (c).

1 Ring Identification and Pattern Recognition

We present an implementation of a ring finding pattern matcher. This particular ring finder is designed to solve what you might call a "join-the-dots" problem, as the locations and the sizes of the rings that it must find are only betrayed by the positions of some dots that have been left at random locations around the circumferences of the rings.

There are many examples of applications for this kind of ring finding pattern recognition.

The particular example that motivated this paper concerned the identification of Cherenkov rings in the LHCb detector. Cherenkov rings are described in more detail in section 1.1, but all we will need to know for the moment is that the size and position of each ring will be betrayed only by the locations of a random number of dots (actually photons) distributed randomly on its circumference.

The single most important thing to recognise when pattern matching is that:

It is impossible to recognise a pattern of any kind until you have an idea of what it is you are looking for.

To give a simple example from the context of ring finding: What rings should a ring finding pattern matcher identify in part (a) of figure 1?

The answer *must* depend on what rings we expect to see!

Equivalently, the answer *must* depend on the process which is believed to have lead to the dots being generated in the first place. If we were to know *without doubt* that the process which generated the rings which generated the dots in (a) is only capable of generating large concentric rings, then only (b) is compatible with (a). If we were know *without doubt* that the process is only capable of making small rings, then (c) is the only valid interpretation. If we know the process could do either, then both (b) and (c) might be valid, though one might be more likely than the other depending on the relative probability of each being generated. Finally, if we were to know that the process only generated *tiny* rings (say only 1/1000th of the width of this page), then there is yet another way of interpreting (a), namely that it represents 12 *tiny* rings of radius too small to see.

So any "join-the-dots" ring finding pattern matcher must incorporate knowledge of the *process* it assumes lead to the production of the dots in the first place.

Inevitably it is impossible to know every detail of the process leading to the generation of the dots, so in practice a ring finding pattern matcher must at the very minimum have a *working model* of the process that leads to the generation of the dots.

1.1 Example of Cherenkov Rings and Cherenkov Photons

When a **CHARGED ???** particle traverses a medium at a speed greater than the speed of light *in that medium*, it emits Cherenkov photons at a fixed **DEPENDING ON WHAT?** angle to its line of flight (but at uniformly random azimuthal angles). In the rest frame of the particle, the allowed region into which these Cherenkov photons radiate is the surface of a forward-swept cone whose tip lies on the particle. With an appropriate optical set-up, it may be arranged that the projection onto a screen of all the cones[1] from a given particle is a ring. The radius of this ring measures the angle at which the photons were radiated. The position of the ring measures the direction in which the particle was travelling through the medium. Because the azimuthal angle of the Cherenkov photons was chosen uniformly, Cherenkov photons will be found uniformly distributed around these rings.[2]

There are many experiments that have imaged (or intend to image) Cherenkov rings in one form or another (**cite LHBC, AMANDA, ICECUBE, SUPERKAMIOKANDE, others**) but each experiment has faced

[1] "All of the cones" here means "not just the cone corresponding to a single instant of time, but the cones corresponding to all times during the particle's traversal of the medium."

[2] Subject to acceptance and optical considerations!

different challenges. Experiments like SuperK have very low event rates and low ring multiplicities and can afford to identify rings with compartitively slow and careful algorithms. Amanda and IceCube lack "projective screens" so they must identify particles via Cherenkov photons picked up at points in three-space. **CITE THEIR RING FINDERS?**

LHCb is faced with a very high event rate, and a very high ring multiplicity (about **XXX** rings per event) and a great deal of overlap between rings. This means that it too needs dedicated ring finding pattern matchers for its own circumstances.

Figure **XXX**BLAH shows a typical collection of Cherenkov photon hits for one event as one might expect to see them in LHCb's RICH-2 detector.

The algorithm that is presented in this paper is optimised for the environment seen in RICH-2 at LHCb. Having said that, the algorithm presented here has a fairly clean separation between those parts of it which are *in*dependent of LHCb and those parts of it which are not.

Later in section **XXX** it will become clear that the innovative parts of the algorithm specific to LHCb are confined to the proposal distribution used by the Metropolis-Hastings sampler. All other parts of the algorithm would be transplantable into any other problem.[3]

The intention of the algorithm is to sample from the space of likely

2 Identifying Cherenkov Rings

2.1 Hits and events in the detector model

In the model, a Cherenkov ring R has a centre $\mathbf{c}$ and a radius ρ. Denote a collection of rings by $\mathbf{R}$.

When a photon is detected by the RICH detector, it is called a hit. For our purposes, the only thing we need to know about the hit is its position h. As we are concerned, an event is characterised by the set $\mathbf{H} = \{\mathbf{h_i} \mid \mathbf{i} = \mathbf{1}, ..., \mathbf{n_h}\}$ of the positions of all of the hits seen in the event.

An example of a a typical collection of hits $\mathbf{H}$ may be seen in figure **XXX**.

The eye is easily able see that, except for a few isolated hits which we will refer to as "background hits", most of the hits in any given event lie on

[3]Note that throughout the *whole* algorithm there is dependence on the *detector model* which is used to measure how likely some set of observed photons would be given an initial distribution of particles. When we say "all other parts of the algorithm would be transplantable into any other problem" we say this assuming detector model were upgraded accordingly.

the circumference of circles[4] as is to be expected from the design of a ring imaging Cherenkov detector.

It is apparent that there is much overlapping of rings in typical events.

2.2 Low-level and high-level event descriptions

An event defined as the set of its hit positions **H** is a low-level description. An event defined as a collection of rings **R** is a high-level description.

2.3 What the `ring-finder` is and is not supposed to do

The purpose of the `ring-finder` is to make statements about likely high-level (ring based) descriptions **R** for an event, given the low-level (hit based) description **H** for that event.

The purpose of the `ring-finder` is **not** to determine the collection of Cherenkov rings that were the cause of the observed collection of it hits **H**.

Rather it is intended that the `ring-finder` should *sample* from the space of all possible high-level (ring based) event descriptions **R** according to how likely they would appear to have been given the observed collection of hits **H**.

Formally, we assume the following:

- We assume that there is a random process M which generates events independently of each other.[5] We assume very little knowledge of M save for the following:
  - We assume that M does not change with time.
  - We assume that in any one event, the mechanism M generates an unknown set of rings $\mathbf{R}^{\mathrm{true}}$ independently of any other events.
  - Where there is reason to believe that the distribution of ring-centres and the distribution of the number of rings generated by M might be known, even if only approximately, we make use of this information.

  We assume that there is a random process P according to which the a collection of observed hits $\mathbf{H}^{\mathrm{obs}}$ is generated from $\mathbf{R}^{\mathrm{true}}$. The random

[4] Sometimes hits lie only on *arcs* of circles due to the finite nature of the sensitive and/or constraints imposed by the design of the optics. In the detector model, only cropping due to the shape of the sensitive area is taken into account. Cropping induced by optical effects is not considered.

[5] M stands for **M**echanism

process P is assumed to be known to a reasonable precision and it may be broken down into parts relating to:

- A uniform distribution of Cherenkov photons about the circumference of the ring,
- A poisson distribution for the number of photons likely to be radiated onto a given ring of radius ρ, and
- Detector resolution.

Formally our detector model may be summarised by a likelihood

$$p_P(\mathbf{H}|\mathbf{R}) \tag{1}$$

and a probability density function

$$p_M(\mathbf{R}). \tag{2}$$

The subscripts M and P on p_M and p_P etc will subsequently be ommitted.

The quantity we are interested in is $p(\mathbf{R}|\mathbf{H})$ which we may obtain from (1) and (2) via Bayes's Theorem:

$$p(\mathbf{R}|\mathbf{H}) = N(\mathbf{H})p(\mathbf{H}|\mathbf{R})p(\mathbf{R}), \tag{3}$$

where $N(\mathbf{H})$ is a normalizing constant which we may ignore (that is arbitrarily set it equal to 1) as we will only be interested in the relative variation of the left hand side of (3) with respect to $\mathbf{R}$ for fixed $\mathbf{H}$. No further mention will be made of $N(\mathbf{H})$.

Note that the dimension of R is three times the number of rings it contains as each ring is defined by a 2-dimensional centre and a 1-dimensional radius. A typical event has contains order 10 rings, and so $p(\mathbf{R}|\mathbf{H})$ is typically a function of order 30 dimensions. It will not therefore be possible to plot $p(\mathbf{R}|\mathbf{H})$. However, we can do the next best thing: we can use montecarlo methods to sample from it.

The set of high-level descriptions R which we will draw from $p(\mathbf{R}|\mathbf{H})$ will represent the most reasonable guesses we can make for $\mathbf{R}^{\text{true}}$ given $\mathbf{H}$.

Note that we do not make any attempt to "maximise" $p(\mathbf{R}|\mathbf{H})$. We are not interested in where this function is a maximum,[6] though in some sense

[6]Note: It may seem strange that we are not interested in the maximum of $p(\mathbf{R}|\mathbf{H})$ given the large amount of the literature devoted to maximum-likelihood analyses. But bear in mind that (1) the position of the peak is not invariant under reparametrisations of the space, and (2) this is a high-dimensional problem, and in high-dimensional problems the vicinity of the maximum is often either only a tiny part of the "typical set" (the region containing most of the mass of the probability) or else is not even part of the typical set at all! Consider **XXX**

the sampling is likely to be localised near the maximum.

2.4 Efficiency considerations

The description of the ring-finding method in the preceding section may be summarised in two steps as follows:

1. Define the types of rings you want to find by calculating the specific forms of the distributions in Equations (1) and (2) which are relevant to your particular rings, and

2. Plot the resulting posterior distribution $p(\mathbf{R}|\mathbf{H})$ (Equation (3)).

Of the two steps above, the first one is by far the simplest. It leaves almost no scope for flexibility or creativity. Either the calculated distributions are or are not a fair representation of the mechanism of ring production, detector response and hit generation for the problem in question. Newer models of the process or the detector can be switched in at a moments notice for comparison with older models, with no significant impact on other parts of the ring finder.

It is the second of the steps above that presents the problems and is the part which will take up most of the rest of the discussion.

It will become clear later that while it would be possible to implement a "general" second step,[7] it is almost certain that such a method would be hopelessly inefficient and completely unusable. We will therefore always discuss the second step in the context of the particular sort of ring finding that is required by LHCb's RICH detector.

Why is a general method of plotting $p(\mathbf{R}|\mathbf{H})$ hopelessly inefficient?

It has already been mentioned that $p(\mathbf{R}|\mathbf{H})$ as a function of R is a function of around 30-dimensions, and we know that we are interested in discovering where in this space the bulk of the probability lies. The large dimensionality of the space precludes a simplistic plot-the-density approach, and suggests that a sampling/explorative method is instead required.

If the space had only one maximum[8] then established techniques (such as steepest descent *etc*) could be employed to find out where the "centre" of the distribution was, and then the simplex or other multi-point methods could probably be used to explore the bulk region.

[7] *i.e.* a second step that is not specifically tailored to the problem in question

[8] **i.e.** if the space had no local maxima other than the global maximum

Unfortunately, the space is actually packed full of local maxima separated by regions of improbability, and so these methods would very quickly get irretrievably stuck in poor local minima.

Take the example shown in Figure 1. There is no way to slowly transform the two fitted rings of (b) into any two of the three rings in (c) which does not involve passing through a huge region inbetween in which the rings would represent a *terrible* fit to the observed hits.

The only real solution appears to be to sample the space using a custom multi-point Markov sampling method – one that has built into it an understanding of the type of space it is trying to explore, and an ability to make sensible guesses as to the locations of distant isolated local maxima.

This is all necessary to improve the *efficiency* of the ring finder to the point at which it can become *useful*. Wherever possible, the choice and design of the custom Markov sampler should not affect the answers that are reached, only the time it takes to reach them.[9]

The approach adopted for the LHCb RICH ring-finder was to use a Metropolis-Hastings sampler with a proposal distribution tailored specifically to the sorts of hits seen in the LHCB RICH detector.

2.5 Custom Metropolis Hastings samplers

A Metropolis Hastings sampler samples a sequence of points $\{r_i\}$ from a target distribution $p(r)$ as follows. Denote the n^{th} point sampled by r_n. A proposed location w_n for the next point is drawn from a "proposal distribution" $Q(w|r)$.[10] The only two requirements of the proposal distribution are (1) that it be easy to draw uncorrelated samples from Q, and (2) that it be possible to calculate $Q(w|r)$ up to an abritrary constant factor. A random number ρ is then drawn uniformly from the interval $[0, 1]$. If ρ is found to be less than $\rho^{\max} = \frac{p(w_n)Q(r_n|w_n)}{p(r_n)Q(w_n|r_n)}$, then the proposal w is accepted, and r_{n+1} is set equal to w_n. Otherwise, r_{n+1} is set equal to r_n.

[9]By way of an example: the simplest possible Markov Chain sampler would probably be one using the Metropolis Method with a flat proposal function. This method would indeed sample the space exactly as required, but you would probably have to wait un unfeasable 100^{30} iterations before the distribution converged on the right answer – as long as a hundred by hundred by hundred scan of all points in the 30-dimensional space would take. The aim of writing a custom sampler is not to change the eventual distribution sampled, or to change the samples themselves, but is only to try to reduce the time needed to arrive at a reasonably reliable samples.

[10]The proposal distribution for $Q(w|r)$ may depend on r, however the "Hastings" part of the name tends to be dropped for symmetric proposal distributions, **i.e.** those for which $Q(w|r) = Q(r|w)$.

Note that a Metropolis-Hastings sampler does not in general produce uncorrelated samples (in fact there is a very high chance that two more more neighbouring samples may actually be identical!) however as the number of samples tends to infinity, the resultant set of samples may be treated as if they were the result of an uncorrelated sampling process.

The art of creating the Metropolis-Hastings sampler that is most suited to a given target distribution is equivalent to finding the best proposal distribution $Q(w|r)$ for the problem.[11] There is considerable scope for creativity in the construction of good proposal distributions for particular problems. Despite much development, the proposal distribution which is being used by the LHCb RICH ring finder at the time of writing (described later **XXX**IN SECTION BLAH) is highly unlikely to be optimal. Further development of the ring finder's proposal distribution is the single most important objective any attempt to improve ring finding performance – judged according to how long you must wait before the samples are representative of the whole distribution.

2.6 The proposal distribution for the Metropolis-Hastings sampler in LHCB's RICH ring finder

The proposal distribution used in LHCB's RICH ring finder is best described algorithmically. At the top level, it can:

1. propose the **addition** of a new ring-hypothesis to the current set of ring-hypotheses,
2. propose the **removal** of a ring-hypothesis from the current set of ring-hypotheses, or
3. propose a small **alteration** to one or more of the existing ring-hypotheses.

By tuning the relative probabilities with which the above options are chosen, one can try to maximise the efficiency of the sampler.[12] A crude attempt at

[11] Note that if $Q(w|r)$ were chosen to be $p(w)$ then $\rho^{\max}$ would always be exactly 1, every proposed point w would thus be accepted, and all samples would be completely independent samples from Q and thus would also be completely independent samples from the target distribution p. This is in some sense the "optimal" Q. But one of the requirements of Q was that we should be able to sample from it, and if we were able to sample from $p(w)$ directly we would have no need of the Metropolis-Hastings method to sample from $p(r)$. So in practice the optimal Q is (a) one from which it *is* possible to draw samples, but (b) is nonetheless "as close as possible" to the target distribution p.

[12] In the context of this document, the efficiency of the sampler is always defined as the reciprocal of the inefficiency of the sampler, which itself is defined as the proportion

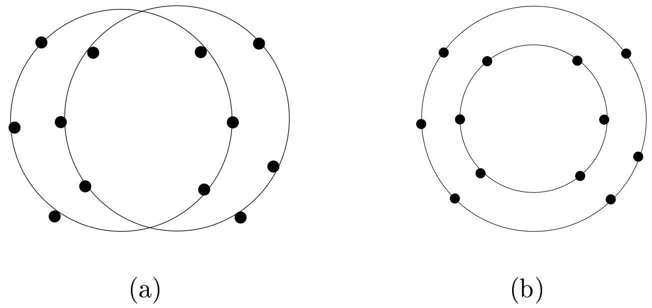

Figure 2: Part (a) shows a common example of a mis-fitting. In this example, the fit should really have been as shown in part (b).

optimisation was done by hand but there will be scope for improvement. At the time of writing, the values in use were $p^{\mathrm{add}} = (1-0.2)*(1-0.4) = 0.48$, $p^{\mathrm{remove}} = (1-0.2)*0.4 = 0.32$ and $p^{\mathrm{alter}} = 0.2$.

2.6.1 Alterations to ring hypotheses

The "alteration" option allows the positions and radii of previously proposed rings to be "fine tuned" to better reflect their likely locations in the light of neighbouring rings etc.

Most of the time, fine tuning is most efficient if it involves perturbing the position and size of only one ring. More often than not, a perturbation to any one ring results in a ring which fits worse than the original, and so as the number of simultaneously perturbed rings grows, the chance that the proposal will be accepted by the Metropolis algorithm diminishes exponentially.

Nevertheless, there is a common situtation depicted in figure 2, in which it is beneficial to try to perturb two rings at once in order to pull a bad fit out of a false minimum. In cases like this, it is unrealistic to expect the ring-finder to switch from 2(a) to 2(b) by successive perturbations of any one ring, or by removal and subsequent reinstatement of both rings, as the potential barrier to this (the poor quality of the intermediate fits) would lead to exceptionally long equilibrium times.

To take this and similar situations into account, the number of ring hypotheses which are the subject of modification in a given 'alteration" is itself chosen at random from a distribution which favours single rings over pairs and pairs over triplets *etc.* The precise choice of this distribution is again

of proposals which are wasted. Wasted proposals are ones which are rejected by the Metropolis algorithm, thus causing the current point to be re-visited as the next point of the sampling.

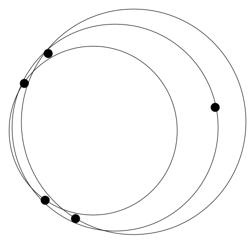

Figure 3: All three of the above rings fit the four points on the left hand side reasonably well. If the “current” ring hypothesis were to be the largest or the smallest ring above, one would like to be able to shrink or grow the ring about the well fitted points until the remaining hit on the right hand side were to also be fitted. No such “correlated smearining” is imlemented in the current code, so the same effect has to be reached by a 3-dimensional random walk, at the cost of efficiency .

something which may be tweaked to increase the efficiency of the sampler. At the time of writing, the number n of ring hypothesis to be altered in a given “alteration” (out of a total number of ring hypotheses N) was selected with a probability proportional to $1/n$. It is likely that a better choice leading to a more efficient sampler could be found.

Once the number n and identity of of ring hypotheses to be perturbed has been chosen, the perturbation of each ring hypothesis is performed by independent symmetrical Gaussian smearings of each of the three ring coordinates (centre-x, centre-y and radius). In each case the width of the smearing is equal to 10% of the average radius of a typical ring in the RICH detector. There is again scope for optimising this mechanism in order to make the sampler more efficient. In particular, when more than one ring is simultaneously modified, it would make sense to allow smearings to correlate between the rings. This might help to more efficiently remove mis-fits like that shown in figure 2. Also it might be an idea to consider correlated smearings within the three parameters of a single ring so that (for example) one could leave the best-fitted parts of the ring as unaltered as possible (see figure 3) while allowing the less well constained parts of the ring to move as much as possible.

2.6.2 Addition of new and removal of old ring hypotheses

Given that the decision to insert-or-remove a ring has been made, the deletion is proposed with probabiity 0.4 and insertion with probability 0.6.

Removal of old ring hypotheses

Once scheduled, removal of a ring hypothesis is as simple as it sounds. The only thing worth mentioning is that the removed hypothesis is not thrown into a black hole and lost forever. Instead it gets pushed onto a stack of "ring hypotheses which have been useful in the past". The use of this stack is discussed later.

Addition of new ring hypotheses

If the three coordinates (centre-x, centre-y and radius) of a ring hypothesis are drawn at random from their whole-experiment average distributions, it is highly unlikely that the resulting ring will correspond to a ring in the data. There may only be 20 to 50 real rings in an event, but there are of the order of $100 * 100 * 20$ distinguishable ring hypothesis you could make. A ring hypothesis drawn at random thus has roughly only a one in ten thousand chance of being close-to-useful. To improve the efficiency of the sampler, proposals for new rings must have a better means of making suggestions.

The approach taken in the RICH ring finder is to try to seed ring suggestions from groups of three hits. Again, it is not good enough to choose just any three hits at random, as a typical event can have upwards of 300 hits (say 15 hits per ring) so the chance of three hits drawn at random coming from the same ring is of the order of $(15/300)^2 = 1/400$ which is still too small to be useful. Instead the three points are chosen in a correlated manner:

The three hit selection method

First one of the hits in the event is chosen at random. Then all other hits in turn are compared with the first hit. For each hit, the likelihood that it (given no other information) is in the same circle as the first hit is calculated. This may be done purely on the basis of the knowledge of the whole-experiment ring radius distribution and a little numerical integration.[13] Once all such likelihoods have been calculated, one of these hits is chosen (with a probability proportional to its likelihood) to join the first. By this stage we have

[13] It might be objected that the whole-experiment ring radius distribution is not known (except from Monte Carlo event generation) before the expermint turns on, and can only be measured in the RICH, and so training the RICH new-ring proposal distribution on the basis of Monte Carlo predictions will introduce some sort of bias into the ring finder. Fortunately this is not a worry, as once again the purpose of the modified proposal distributions is not to change the answer,[14] only to reach the answer more efficiently. If the Monte Carlo data were not to match the experimental data very well, that would only make this proposal distribution a little bit more inefficient than intended ... it would not invalidate the result otherwise.

selected two hits which have a reasonable probability of being in the same ring. We now need to choose a third. A similar procedure is followed as before. All other hits are compared with the first two, and the likelihood that (given no other information) they are in the same ring as the first two is calculated, and one of the hits is then chosen to join the first two with a probility proportional to its likelihood of being in the same ring. Again this depends only on knowledge of the whole-experiment ring radius distribution. Having selected three points likely to be in the same ring, the ring passing through all three hits becomes the proposal which is offered to the Metropolis method for approval or rejection.

Of course, badly distorted rings and rings with fewer than three hits on them will never be seeded this way, so the above prescription is applied only 90% of the time. The remaining 10% of the time the proposal function falls back on a naughty trick – it suggests a "reverse ring addition". A "reverse ring addition" is not strictly a valid action in the context of the Metropolis method as it breaks the principle of detailed balance. However in practice it has proved to be a good thing to have inside the sampler, and it does not seem to break the principle of detailed balance enough to cause any obvious problems. All a "reverse ring addition" is, is the popping off and subsequent re-use of the top-most ring in the stack of "ring hypotheses which have been useful in the past"[15] as the new-ring proposal. In the event that the stack is empty, the proposal method falls back to the most basic of the new ring proposal mechanisms described at the beginning of this section, even though it is very unlikely to lead to much success. It appears that the "reverse ring addition" mechanism allows the ring-rinder to be a little more aggressive about throwing ring hypotheses away than it would otherwise be able to be – if it gets second thoughts about a disposal, it effectively has a chance to change its mind and recover a ring hypothesis that it had thrown away earlier.

[15]See "Removal of old ring hypotheses" at the start of this section

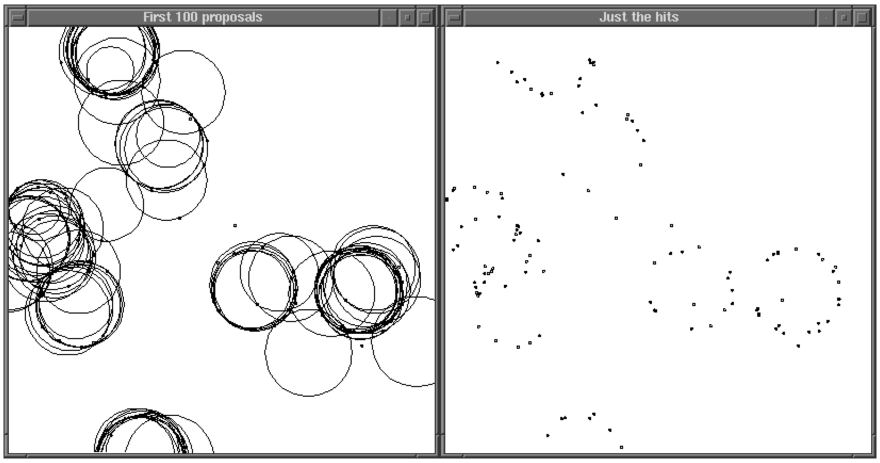

Figure 4: a (left) 'original'`HERWIG`, b (right) `ISAJET`