Investigating Emergence by Coarse Graining Elementary Cellular Automata

Andrew Weeks[1], Fiona Polack[1] and Susan Stepney[1]

[1]Department of Computer Science, University of York, UK, YO10 5DD

Abstract

We extend coarse graining of cellular automata to investigate aspects of emergence. From the total coarse graining approach introduced by Israeli and Goldenfeld, *Coarse-graining of cellular automata, emergence, and the predictability of complex systems*, Phys. Rev. E, 2006, we devise partial coarse graining, and show qualitative differences in the results of total and partial coarse graining. Mutual information is used to show objectively how coarse grainings are related to the identification of emergent structure. We show that some valid coarse grainings have high mutual information, and are thus good at identifying and predicting emergent structures. We also show that the mapping from lower to emergent levels crucially affects the quality emergence.

Introduction

We are interested in observing and modelling complex emergent systems, with the goal of understanding how we could begin to specify and implement engineered emergent systems. Emergence is variously characterised; we start from Ronald et al's definition of emergence: "The language of design L1 and the language of observation L2 are distinct, and the causal link between the elementary interactions programmed in L1 and the behaviors observed in L2 is non-obvious to the observer..." (Ronald et al., 1999). Here, we refer to the local level, of the implementation substrate, as L. The language of observation represents a global, or coarse-grain, level where emergent behaviour is observable that we refer to as the specification, S. After Shalizi (2001), we define emergence in information-theoretic terms, as the greater *predictive efficiency* of descriptions in S over those in L.

In natural complex systems, it is hard to define languages L and S, and to determine accurate mappings between them. Here, the complex emergent systems are elementary cellular automata (ECAs); their language is simple and well-defined, and thus mappings can be identified and analysed.

One perception of an emergent system is that its high level behaviour is independent of the low level behaviour. However, the emergent properties are actually a carefully chosen extract of the low-level behaviour. The observational discontinuity allows us to identify emergent behaviour. Elsewhere (Weeks et al., 2007), we show that coarse graining is a simple form of emergence. If we can state coarse-grained rules, then we can use the coarse level to predict behaviour. Because information is lost in the higher level we cannot predict behaviour correctly in all cases, but the rules should be able to predict some common futures. Here, we explore emergence through coarse graining ECAs and measurement of mutual information between levels.

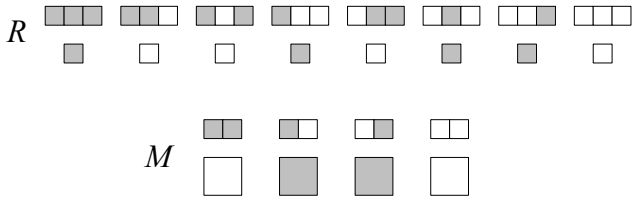

Figure 1: Rules and mappings. Shaded cells represent value 1. R shows how the "name" of an ECA rule is derived: each of the 8 possible ECA initial state is shown in bigendian order; below it, the next state of the central cell is shown; the rule is "named" by reading off the values of the next state. Here, R is the transition table for ECA rule 150 (10010110_2). M represents the coarse grain mapping 0110, with grain $g = 2$.

Coarse Graining ECAs

An ECA is a one-dimensional cellular automaton, with two states and a neighbourhood of three. There are 256 ECA rules, of which 88 are distinct (not just spatial reflections or 0-1 inversions). Rule sets are named by taking the decimal representation of the binary string that represents the outputs of the transition rules from all neighbourhood states taken in bigendian order (figure 1).

The coarse graining of ECAs was investigated by Israeli and Goldenfeld (2006). In a coarse graining at grain g, the values of a contiguous block of g cells at the fine level are projected, or mapped, to the value of a single cell at a coarse level (figure 1).

Israeli and Goldenfeld (2006) require their coarse grainings to be *total*, that is, to satisfy the commutativity condition that running the fine ECA for $n \times g$ time-steps then

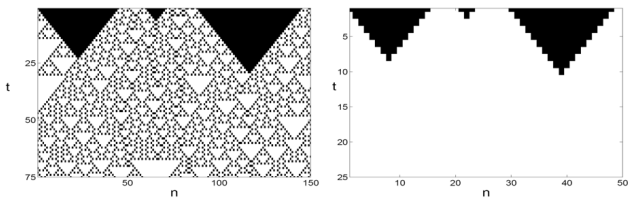

Figure 2: Space-time plot of 75 time-steps of rule 146, and its total coarse graining to rule 128, with grain 3, from Israeli and Goldenfeld (2006).

performing the mapping, gives the same result as performing the mapping, then running the coarse ECA for n time-steps. We relax this condition, to discover *partial* coarse grainings that, while not 'correct' in this commutative sense, nevertheless have good predictive properties (high mutual information). We do not consider 'trivial', and thus uninteresting, mappings, that map either to all zeros or to all ones.

Essentially, the coarse graining represents a system specification (language S) for the ECA at the fine level (language L). In one sense, the languages S and L differ only in the grain of the representation – both languages are the language of rules of ECAs. However, we could also take the language to be the specific ECA rule. In either case, ECA coarse graining reduces the language mapping to a tractable problem, and provides a starting point for exploring emergent behaviour.

Israeli and Goldenfeld (2006) show that almost all ECA rules can be coarse grained: their behaviour is mimicked exactly at a coarser grain by other ECA rules (figure 2). Note that the mapping is applied only in the initial state, to set up a correspondence between the fine and coarse grained initial state. Subsequently each ECA runs independently: the validity of the coarse graining ensures that the mapping always holds from then on.

Because of the exact mapping between the fine and coarse grained ECAs, the coarse grained ECA consistently predicts aspects of the state of the fine-grained ECA for any point in the future (since information is lost in the coarse graining, the prediction is not absolute, but it is consistent). For $g = 2$, running the coarse grained ECA requires only 25% of the calculations of the associated fine-grained ECA (assuming we calculate the next states naïvely).

As with any emergent system, one aim of coarse graining is to end up with a compact representation of the high-level behaviour of the underlying system (or some aspect of it). In the case of coarse graining, that compact representation takes the form of another ECA rule (but one that operates at a coarser grain). Clearly the high level model will predict only certain aspects of the system.

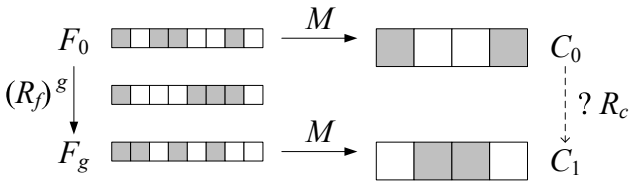

Figure 3: Steps in discovering a coarse graining, $g = 2$, illustrated for a 6-cell initial string. As in Israeli and Goldenfeld (2006), each non-overlapping block of g cells in a fine F state is mapped to one cell in the coarse C state. ECA rule R_f evolves through the three steps on the left, F_0 to F_g. State F_0 is mapped to state C_0 using the mapping 0110. The same mapping is used to produce C_g from F_g. When these steps are followed for a complete initial state (384 cells, see step 1), the candidate coarse grain rule, R_c can be read off, as in step 5, below.

Finding a Total Coarse Graining

Finding coarse grainings is a systematic process. Candidate mappings are successively proposed and applied; the result of each mapping is checked to determine whether the mapping generates a consistent coarse rule. We describe an algorithm that can be used to find all the total coarse grainings of an ECA rule R_f at grain g with non-trivial mapping M. Like Israeli and Goldenfeld (2006), we use the same grain g for the cells (spatial) and the time-steps (temporal). This maintains the speed of information propagation – for instance, if the spatial grain is less than the temporal grain, then information propagates too fast for ECA rule capture.

The application of steps 1 to 5 of the total coarse graining algorithm is illustrated in figure 3.

1. Construct the initial state for the fine-grained ECA, F_0. For a total coarse graining, we must guarantee that the coarse grained version of the initial state contains all eight (000, 001 ... 111) neighbourhood states. This will allow the coarse ECA rule to be read off in step 5. It is sufficient to include all possible states of $3 \times g$ cells (where 3 is the neighbourhood size of an ECA) in the fine-grained initial state, giving a string of length $2^6 \times 6 = 384$ for $g = 2$.

2. Run the fine-grained ECA for the equivalent of one coarse time step: g time-steps at the fine grain, resulting in fine-grained state F_g. We now have the underlying fine states for two successive time steps of the coarse CA.

3. Apply the mapping M to the initial state of the fine-grained ECA, F_0, to produce the initial state of a candidate coarse-grained ECA, C_0. Apply M to the final state, F_g, to produce the next state of a candidate coarse-grained ECA, C_1.

4. C_0 and C_1 are not necessarily related by an ECA rule. For each of the eight possible neighbourhood states σ_i,

check that every instance of σ_i in C_0 maps to the same value in C_1. (It is sufficient to consider only distinct triplets of cells in C_0: if these give consistent states in C_1, then all the overlapping neighbourhoods do so too, by construction of the initial state in step 1.)

5. If the coarse states C_0 and C_1 are consistent with successive states under some ECA rule, then 'read off' that rule R_c, by locating the eight neighbourhood states (111, 110 ...000) in C_0, and recording the values they map to in C_1 (as shown for R in figure 1). (Because of the consistency check, these values are unambiguous. Because of the construction of F_0 and the non-triviality of M, it is always possible to locate at least one instance of each of the eight neighbourhood states in C_0, so the rule is totally defined.)

The number of non-trivial mappings M is $2^{2^g} - 2$, which is low for small g (14 for $g = 2$), so an exhaustive search over mappings is efficient. (Even so, it is not guaranteed that a coarse graining of a rule at a particular granularity exists with any mapping.) At higher g, other factors render the discovery of coarse graining intractable, before the effect of the number of mappings becomes intractable.

It would also be possible to perform a coarse graining by stating the fine and coarse ECA rules and calculating a mapping, but this is a less efficient approach, with more consistency checks to perform.

For the 256 ECA rules, there are 182 non-trivial total coarse graining relationships at $g = 2$. For the 88 unique and non-trivial ECA rules, there are 35 non-trivial total coarse grainings.

Extending Coarse Graining to Partial Mappings

Information must be lost when coarse graining. Sometimes the fine detail of the original rule disappears: Israeli and Goldenfeld (2006) refer to this as a loss of irrelevant degrees of freedom (DOF). In other cases relevant DOF (Israeli and Goldenfeld, 2006) are lost, meaning that the information that is propagated by the fine CA cannot be modelled in all cases by its coarse counterpart. A total coarse graining precisely captures the relevant aspects of the underlying dynamics, losing only detail that is irrelevant at the coarse level. (Also, where detail that is relevant to propagation at the fine level is lost, it is unimportant, because propagation to the coarse state is unaffected.) In this sense, total coarse grainings are simply compressions of the fine ECA rule.

Israeli and Goldenfeld (2006)'s approach requires coarse grainings to be *total*; that is, enough information must be available in state C_0 for the coarse rule R_c to be read off unambiguously, with the consequence that the fine and coarse ECAs evolve consistently. If we relax the totality requirement, we can provide an initial fine-grained state that does not cover all possible input conditions, and deduce a set of *partial* coarse grainings. A partial coarse graining results in a coarse rule R_c that does not necessarily reflect the underlying rule R_f in all cases. Even so, this relaxation can allow the coarse grained rule to approximate more of the underlying behaviour than would otherwise be possible. An ideal partial coarse graining is one for which the initial conditions admit a broad range of the fine rule's behaviours, reflecting the features of most interest. An analogy can be drawn with physical emergent systems, where the emergent properties occur over a restricted set of all possible low-level states, such as a certain temperature range.

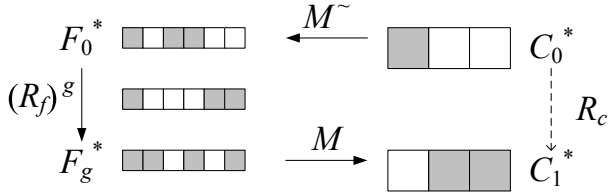

Figure 4: The extra steps involved in discovering a partial coarse graining (see text for details).

It may be thought that total coarse grainings are ideal, and we should aim to get partial coarse grainings as close to that as possible. While broadly true, being total is neither necessary nor sufficient for a good coarse graining – for the discovery of a coarse ECA rule that models the desired high level behaviour (we elaborate measures of goodness below). It is easy to see that some total coarse grainings capture uninteresting aspects of the underlying fine-grained rule; for example, even with non-trivial mappings, many rules coarse grain to rule 0 or rule 255; these are valid coarse grainings that convey no information about the underlying behaviour. In applying partial coarse grainings, we seek to find coarse rules that capture the maximum of useful (predictive) behaviours from the underlying rules, at the expense of allowing the coarse graining to make occasional mistakes.

Finding a Partial Coarse Graining

The approach for partial coarse graining follows almost the same steps is as the total coarse graining. However, when F_0 is constructed in step 1, it does not include all possible states of $3 \times g$ cells, and thus the initial state can be smaller than for total coarse graining. Consequently, to extract as much information as possible, steps 4 and 5 consider every (overlapping) combination of three cells in C_0.

Because the constraint on F_0 has been relaxed, C_0 may not include every possible rule case, meaning that the coarse rule cannot be read off unambiguously in step 5. To complete the rule set, we can add any rule case that gives a consistent result in the coarse graining. This can be derived as follows (see also figure 4):

5a. Create a coarse state C_0^*, comprising the neighbourhood

states missing from the coarse state C_0,

5b. Use the mapping in reverse, $M^{\sim}$, on C_0^* to create a corresponding F_0^*. This reverse mapping is a relation; any of its restrictions to a function can be used to create F_0^*.

5c. From F_0^*, run the fine ECA rule for g time-steps to create F_g^*.

5d. Apply M to F_g^* to produce C_1^*.

5e. Apply step 5 to C_0^* and C_1^*, thus reading off rule cases to total the coarse ECA rule set.

We have investigated various approaches to selecting an inverse mapping (step 5b). The most conservative results come from constructing all the inverse mappings, but accepting only those that are totally consistent. Since partial coarse graining is based on a proper subset of possible mapping relations, completions of the coarse rule set result in more valid coarse grainings of each rule than total coarse grainings. For the conservative approach to completion, and a grain $g = 2$, it is not unusual to get 50% more partial coarse grainings than total coarse grainings.

Exploring Partial Coarse Grainings

To assess the parameters of partial coarse graining more fully, we conduct a variety of experiments on ECA coarse grainings. For $g = 2$, we select different initial states, and find partial coarse grainings, using the conservative completion above. As the length of the initial state is increased, and the number of cases that it covers increases, there is a fall in the number of coarse grainings to valid ECA rules. In our experiments, a 6-bit input string (1.5% the length of the complete initial state used for total coarse graining) produced 57% more (322 : 182) coarse grainings than the total coarse graining under the same mappings. This is to be expected, since a short initial state is created by concatenation of only a small proportion of all possible states, so there are potentially many missing cases in the candidate coarse rule, with the potential for several different coarse rules to be derived from different completions of the rule set. However, the partial coarse grainings of a rule include all the rules to which there are total coarse grainings in each mapping, thus we can say that very short initial strings (and thus reduced calculation load) produce results that are consistent with those from the much longer initial state of a total coarse grainings.

There is a limit to the shortness of an initial state string. Here, this is set by features of the algorithm used for coarse graining. An initial state of fewer than 6 bits is not efficient, as it results in interference due to the wrapping of the state (periodic boundary conditions).

Our experiments also show that the form of a short initial state string appears to have a marked effect on the quality of partial coarse grainings obtained. The initial state string

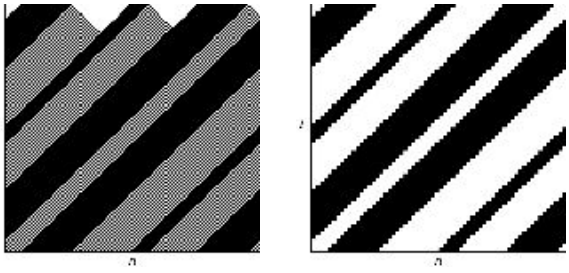

Figure 5: Space-time plot of 50 time-steps of rule 186, and its partial coarse graining to rule 170, with $g = 2$

101010, for instance, produces fewer partial coarse grainings to valid rules at $g = 2$ than the string 101101, but the partial coarse grainings from the string 101010 are judged to be of better quality as predictors of structure than those from string 101101. We cannot yet generalise from these results, not least because, at $g = 3$, these two initial strings produce very similar results. Furthermore, the initial state string 101010101101, made up of both these elements, produces good quality results at both $g = 2$ and $g = 3$. We are still investigating ways to determine what makes a good short initial state, and to determine how the quality of coarse grainings (or proportion of good coarse grainings) might relate to initial state.

We conclude that, although partial coarse grainings do not provide total accuracy in their predictions of fine grained behaviour, they can still provide 'good' descriptions. Figure 5 gives such an example: a partial coarse graining of rule 186 to rule 170. The coarse graining captures the significant, and persistent, "diagonal" structure of the fine rule. Note that under total coarse graining at $g = 2$, rule 186 coarse grains to rule 128, capturing only the transient "triangular" structure of the fine rule.

Intuitively, we can see that the capture of persistent structure by the partial coarse graining is more significant than the capture of the initial transient behaviour by the total coarse graining, but we would like an objective measure of this 'goodness'. In the next section, we show how to quantify what we mean by 'good' in information theoretic terms.

Quantifying Emergence

A challenge of emergent systems engineering is to be able to determine which low-level system gives a good emergent behaviour. In studying coarse graining, we want to be able to distinguish, as objectively as possible, good coarse grainings (that capture interesting properties of fine-grained ECAs) from other legitimate coarse grainings. Here, we use an information theoretic measure of mutual information.

Information theoretic measures of emergence have been proposed by Crutchfield (1994), and more recently by Shalizi (2001), Prokopenko et al. (2007), and others. It has been shown that the mutual information I between the implemen-

tation level L and the observational level S of a system is a measure of emergence. The intuition is that I measures the amount of information in a low level model (in a language L) that can be predicted by a higher level model (in language S). Modelling, or incremental system development, can be viewed as increasing the shared information between the specification and implementation. In the case of coarse graining, a good mapping can be thought of as a mapping that results in a high I between the fine-grained ECA (L) and the coarse ECA (S).

Mutual information can be calculated using a suitable entropy measure H of the systems, and can be expressed either in terms of the joint entropy $H(S, L)$ or, equivalently, in terms of the conditional entropy $H(S \mid L)$, or $H(L \mid S)$ of the systems.

$$\begin{aligned} I(S;L) &= H(S) + H(L) - H(S,L) \\ &= H(S) - H(S \mid L) \equiv H(L) - H(L \mid S) \end{aligned} \quad (1)$$

Intuitively, $I(S;L)$ is the correlation between the specified system (the coarse-grained ECA) and its implementation (the fine-grained ECA). In terms of conditional entropy, $H(S \mid L)$ is the information in the system specification that is not captured by its implementation, whilst $H(L \mid S)$ represents properties of L that do not explain, in information theoretic terms, the observed properties of S.

Calculating Mutual Information for ECAs

Mutual information requires a suitable entropy measure. We need an efficient entropy measure for ECAs, and, for calculating the I, we need to take account of the different spatial and temporal scales at the fine and coarse levels. Other research (for instance, (Zhao and Billings, 2006; Mori et al., 1998)) has used I in relation to ECAs, but with different purposes; their measures, though similar, do not directly adapt to our requirements in relation to spatial and temporal scales.

The key to a meaningful entropy measure is to measure over an appropriate scale, so that it identifies the structures that are important at that scale, without too much influence from order at other scales. This can be seen if we consider the entropy of a system such as a flocking simulation. We could measure the entropy at the level of each individual, in which case the entropy rises as individuals form flocks, because it is harder to characterise the behaviour of an individual in a flock (velocity, position, flock influences) than it is to characterise the behaviour of an isolated individual (velocity, position). Alternatively, we can attempt to measure the entropy of groups of individuals, in which case the entropy of a group that forms a flock is lower than that of a group that is incoherent, because it is easier to describe the flock's behaviour than that of an incoherent group of individuals.

Turning to ECAs, the rate of information transfer is limited by the neighbourhood, and a neighbourhood value represents one input in the initial state of an ECA; thus it seems reasonable to consider entropy in relation to the neighbourhood. Entropy is calculated from the probabilities of occurrence of each possible neighbourhood chunk value (000, 001, 010 ...), measured over many runs of each ECA. As we need to calculate I between two ECAs, we use the coarse grain for chunking, so entropy is measured over fixed chunks of three coarse cells. Therefore, the fine grain entropy, at $g = 2$, is calculated over chunks of six fine cells. Note that this means that fine-grained ECAs have a higher maximum entropy than coarse ECAs.

Accounting for the temporal difference between the coarse and fine grains requires identification of the generations of each ECA to be measured. After analysis of the practicalities of calculation and a number of experimental investigations, we determined that every coarse grained generation (row) should be measured, but only the corresponding fine-grained generation – for $g = 2$, that means every alternate fine-grained row. This means that the entropy of the fine grained rule overlooks some of the behaviour of the fine ECA, notably that over short time-scales. This would be a problem if we were interested only in entropy as a measure of complexity; the effects of the contraction are still being considered in our analyses.

The algorithm that we use to calculate entropy starts with the selection of a set of initial state strings on which to base the calculations. We calculate I for each coarse graining pair of ECA rules, and for each mapping that defines this coarse graining. Firstly, for one rule pair and one mapping, each test string is used as the initial state for the fine-grained ECA; the mapping is used to derive the equivalent coarse grained initial state, and the two ECAs are then run. We then select the rows representing the generations of interest, all coarse grained rows and the equivalent (gth) fine grained row, and apply the chunking equivalently to each pair of rows. We can then identify the number of times that each chunk occurs in each row.

These measurements are done for all the chosen initial state strings, and are then used to calculate the probabilities of the different chunks; the probabilities are used in the usual Shannon entropy calculations, yielding entropies $H(L)$ (from the fine grain), $H(S)$ (from the coarse grain) and $H(S, L)$. From these, I is calculated (equation 1).

We have completed these calculations for all $g = 2$ partial and total coarse grainings, and for a selection of coarse grainings at other granularities. We have also looked at I calculations based on different chunkings and time divisions, and, to date, this approach gives the most useful and cost-effective I measures.

Mutual Information and Coarse Grained ECAs

Mutual information can be a useful guide to the goodness of an emergent solution, and, here, an indicator of the quality of a coarse graining. A good mapping results in a high I

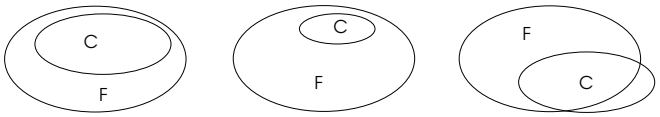

Figure 6: Representation of three different coarse grainings. The ovals indicate the amount of entropy: the large ovals represent the entropy in some fine grained ECA, F; the smaller ovals the entropy in the corresponding coarse grained ECA, C; the overlap represents the mutual information. The left figure shows a total coarse graining with high I; the middle a total coarse graining with low I; the right a partial coarse graining with high I.

between the ECAs. Here, we consider factors that influence I of mappings between ECA rules in coarse grainings.

I is high if the ECAs' behaviour is non-trivial (complex or chaotic rules) and tightly coupled (they mirror each other closely). The ECAs in a total coarse graining must always mirror each other's behaviour, so I is always maximal: the mutual information is exactly the entropy of the coarse ECA rule. However, if the behaviour of the ECA rules is trivial, and thus entropy is low, the maximal I for that coarse graining is lower than it is for "interesting" ECA rules. Furthermore, a non-trivial fine ECA rule (with high entropy) could be coarse grained to a simple rule (with low entropy), and the maximal I for the total coarse graining must then be low for that mapping. This is illustrated in figure 6. Total coarse grainings are accurate, but they are not necessarily good.

We have measured the I of total and partially coarse grained ECAs. As expected, many coarse grainings (both total and partial) have low mutual information – the coarse graining is not highlighting any significant structure from the fine grained ECA. This can occur even when the fine grained ECA rule produces non-trivial structure, with the extreme case being coarse grainings from a complex fine grained rule to a vacuous coarse grained rule such as rule 0 or rule 255. Furthermore, the additional coarse grainings that are valid under partial coarse graining include many vacuous cases with low I.

However, one result in particular is exciting in terms of the potential for using coarse graining to predict emergent structure – that high-I partial coarse grainings seem to predict high-I total coarse grainings at the current and higher grains. If we study all the total coarse grainings at a particular value of g, we find that some rules are coarse grained by many rules, and that the coarse grainings have notably high I – intuitively, this would suggest that there are structures in the fine-grained rule that are (a) non-trivial and (b) common to many rules at the coarse grain. When we consider the same results for partial coarse graining, firstly we observe that all the total coarse grainings are found by the partial approach, and that partial coarse grainings from rules that are associated with high-I total coarse grainings usually have higher I than the total coarse grainings – partial coarse grainings tend to predict more of the fine-grained structure than total coarse grainings. Next, we observe that the *additional* partial coarse grainings that have high I tend to be those that link rules with many total coarse grainings. It is also the case that these partial coarse grainings have higher I than partial coarse grainings that do not form links among rules with good total coarse grainings. Intuitively, where there is structure to exploit, the partial coarse grainings exploit more of the structure of the fine rule than the total coarse grainings, and do so preferentially where there are already good total coarse grainings.

Furthermore, where we observe high-I partial coarse grainings, these are good predictors of total coarse grainings at higher granularities, and in particular of non-trivial total coarse grainings (good partial coarse grainings at $g = 2$ predict total coarse grainings at higher g).

We note that we have measured I both very accurately and approximately. Tests with sufficient data to give good statistics (computationally rather expensive) were taken by averaging over 50 runs with different random initial states of 1000 characters and equal probabilities of the two cell values. An example of an approximate measurement is a single running of a 384-character initial state, with equal probabilities (for $g = 2$ this is the smallest complete state). For all mappings and all rules, the approximate I results are close to the results of the same extensive tests. Thus, we get good estimates of I, and hence of the 'goodness' of the coarse graining, more cheaply.

The Importance of the Mapping and Timing

We have noted elsewhere (Polack et al., 2005) that the mapping between the implementation and the observed system is essential to the construction of an emergent system. Without a good mapping it would be impossible to use a high level model (even an otherwise valuable one) to predict system behaviour. The work on ECA coarse grainings shows that, even where languages are essentially the same at two levels, a good mapping is hard to systematically derive, and is not always natural or obvious.

In fact, it is often the case that a fine rule R_f can be coarse grained to the *same* coarse rule R_c by different mappings M and M'. Figure 7 shows coarse grainings of fine rule 160 (the top row – the two columns represent two different initial states) to coarse rule 128 with two different mappings (second and third rows). Furthermore, some rules coarse grain to themselves with different mappings: for $g = 2$, rule 150 partially-coarse-grains to itself with 14 different mappings, of which six mappings (those with equal numbers of 0s and 1s) are also total coarse grainings. In general, mutual total coarse grainings with different mappings have similar Is, whilst for partial coarse grainings, some of the mappings have significantly lower Is than others; again rule 150 is

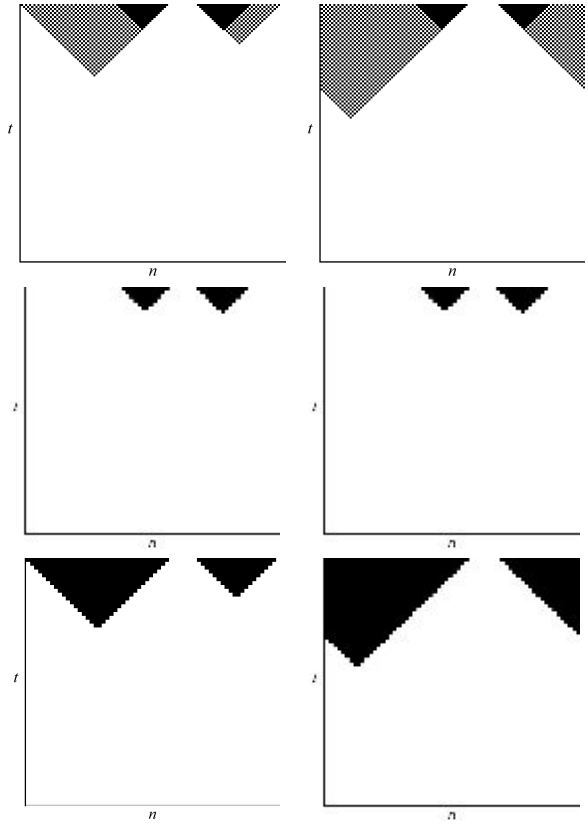

Figure 7: Space-time plots of coarse grainings of rule 160 to rule 128. The top diagrams are rule 160 with two different starting states. The next two diagrams are the results of coarse graining to rule 128 with the mapping 0001. The last two diagrams are the results of coarse graining to rule 128 with the mapping 0101. All figures represent 50 fine time-steps with $g = 2$

an extreme case, with total mappings to itself having reasonable predictive power (I values of about 3), whilst the partial coarse grainings have low predictive power (Is are around 0.6).

One important factor in prediction, that has a marked effect on measures of mutual information, is detection of transient or longer-term features of the ECA. In figure 7, we see that these ECAs become quiescent after about 20 fine time-steps; in the first coarse graining, quiescence is reached much sooner. Mutual information would be different if measured whilst transient behaviour dominates, compared to post-transient. The same is true for rules that have steady-state behaviours after transients die out, such as rules 186 and 170 (figure 5). Table 1 gives an example of mutual information and entropy measures for two partial coarse grainings of rule 162, measured at 4 and 10 coarse time-steps. Coarse graining to rule 128 (mapping 0001) has a similar short-term transient behaviour to that shown in figure 7, whereas the coarse graining to rule 170 (mapping 0111)

C	t	$H(162)$	$H(C)$	$I(C;162)$
170	4	4.354	2.386	1.879
	10	4.293	2.405	1.825
128	4	4.354	0.276	0.276
	10	4.293	0.093	0.093

Table 1: The effect of time-step on mutual information of coarse grainings of ECA rule 162, at $g = 2$. The start state has 1000 cells, and entropy is calculated over 50 runs.

picks up persistent diagonal features.

Coarse rule 170 shows only a slight (non-significant) change in mutual information between time-steps, which shows that rule 170 captures some persistent behaviour in rule 162. For coarse rule 128, however, there is a significant change in mutual information between time-steps 4 and 10. In most of the 50 runs, rule 128 reaches quiescence by coarse step 10 (no information, so no mutual information). Note that the mutual information data show that the partial coarse graining to rule 128 is also a total coarse graining – the mutual information is the same as the entropy of the coarse rule.

In most work on emergence, the focus is on the behaviours and languages at the high and low levels. The selection of mappings, and of total or partial coarse graining, is the subject of ongoing research; however, our work shows that the mapping between high and low levels is an important component of the emergence. This can be interpreted as the way the low level system is viewed through the mapping to form the high level (emergent) description.

Discussion

Total coarse graining of an ECA at $g = 2$ is efficient and fast. However, because of the number of calculations and checks to be performed, $g = 3$ is exponentially slower, and we found that $g = 5$ is beyond the limit of capability of a desktop computer. Partial coarse graining provides a tractable alternative, because high-I mappings can be determined from a small initial state. Furthermore, because the low-g partial coarse grainings are good predictors of higher-grain total coarse grainings, higher granularity searches can be focused rather than exhaustive.

One feature common in partial-only coarse grainings is the ability to predict beyond the transient behaviour of the rules. A total coarse graining often predicts early behaviour accurately but then dies out to quiescence – the total mapping over-constrains the ability to predict the long-term behaviour (as in figure 2). However, a partial coarse graining is less constrained, and is free to mismatch some early behaviour; thus it may be able to predict long-term behaviour (as in figure 5).

Despite the implication of the name "emergence", in information terms, emergence does not add anything; it re-

moves (or hides) aspects of the underlying system, to emphasise an apparently coherent core behaviour which we identify as a higher level phenomenon of interest. In effect, our high level view (constrained by language, time, etc.) blurs the underlying system so that only certain aspects of its behaviour are apparent. This is precisely equivalent to the DOF lost when coarse-graining an ECA.

Our work uses an information theoretic measure to compare the emergent quality of coarse grainings, but it does not explain how the measure can be maximised by selection of a good mapping rule, nor does it tell us how to relate information-theoretic measures to subjective qualities of desired (or deprecated) emergence.

Information theoretic approaches, such as comparison of I used here, have the potential to help determine a good solution, so long as we can map the desirable properties to information theoretic features. We have also found that a limited (and quick) I test approximates closely the results of an equivalent extensive (statistically valid) I test. This is important – if we consider that any valid coarse graining is identifying something emergent, then we have shown that finding emergence is easy; however it is the analysis of I that distinguishes useful (structure-finding) emergence from vacuous or trivial emergence.

We have found that a surprisingly small initial state string predicts results almost as well as using a complete initial state string, and moreover that the extra rules mapped by such a partial coarse graining are themselves useful indicators of interesting properties at coarse grains. If we think of the granularity as the scale of the emergent property, this is hinting at relationships across several scales, and hence across several levels of emergence.

We have demonstrated the importance of the mapping between the levels. Correctness is not necessarily an indication of goodness: having a mapping and a rule that works is not always enough. Indeed, it is often not enough (the result can have low I). Finding a valid coarse graining is much easier than finding a good one. It is, however, the extra 'goodness' properties that can be exploited to get robust implementations of S.

Our immediate future work is to investigate the relationship between partial fine state string and goodness of results (why do such short strings work so well here, and do they always?), and establish the relationship between 'good' coarse grainings at grain g and total coarse grainings at higher g. It is clear that different mappings and grainings focus attention on both different qualities of behaviour and different durations (transient, persistent) in the fine rule. It may ultimately be possible to tailor the selection according to what is of interest at the time.

Whilst coarse graining has a number of interesting properties, our main aim in this work is to gain understandings of these simple emergent systems that will allow us to progress two larger research goals: to understand and engineer emergence, and to find efficient solutions to difficult problems. Coarse graining is an efficient way to detect concealed structures, and thus might be applicable to guided search techniques. More importantly, our work shows that finding and exploiting mappings is likely to yield further progress, in guiding search for solutions.

Acknowledgements

Andrew Weeks is funded by a Microsoft Research European PhD Scholarship.

References

Crutchfield, J. P. (1994). The calculi of emergence: Computation, dynamics, and induction. *Physica D*, 75:11–54.

Israeli, N. and Goldenfeld, N. (2006). Coarse-graining of cellular automata, emergence, and the predictability of complex systems. *Phys. Rev. E*, 73:026203.

Mori, T., Kudo, K., Namagawa, Y., Nakamura, R., Yamakawa, O., Suzuki, H., and Uesugi, T. (1998). Edge of chaos in rule-changing cellular automata. *Physica D*, 116:275–282.

Polack, F., Stepney, S., Turner, H., Welch, P., and Barnes, F. (2005). An architecture for modelling emergence in CA-like systems. In *ECAL 2005*, volume 3630 of *LNAI*, pages 433–442. Springer.

Prokopenko, M., Boschetti, F., and Ryan, A. (2007). An information-theoretic primer on complexity, self-organisation and emergence. *to appear*. available at http://www.ict.csiro.au/staff/mikhail.prokopenko-/Publications/Agents/ITprimer-Nov-2007.pdf.

Ronald, E. M. A., Sipper, M., and Capcarrère, M. S. (1999). Design, observation, surprise! a test of emergence. *Artificial Life*, 5(3):225–239.

Shalizi, C. R. (2001). *Causal architecture, complexity and self-organization in time series and cellular automata*. PhD thesis, University of Wisconsin at Madison.

Weeks, A., Stepney, S., and Polack, F. (2007). Neutral emergence and coarse graining. In *ECAL 2007*, volume 4648 of *LNAI*, pages 1131–1140. Springer.

Zhao, Y. and Billings, S. A. (2006). Neighborhood detection using mutual information for the identification of cellular automata. *IEEE Transactions On Systems, Man, And Cybernetics Part B: Cybernetics*, 36(2):473–479.