A NOTE ON MINIMAL COVERS FOR SOFIC SYSTEMS[1]

MIKE BOYLE, BRUCE KITCHENS AND BRIAN MARCUS

ABSTRACT. We characterize the sofic systems which have minimal subshift-of-finite-type covers and derive some consequences.

0. Introduction. For relevant background, see §1. By a factor map, we mean a continuous onto shift commuting map between two symbolic systems. By definition, a sofic system S is a symbolic system which is a factor of a shift of finite type (SFT). In the language of automata theory, sofic systems are regular languages. A *minimal cover* for S is a SFT, Σ_A, and a factor map $\pi: \Sigma_A \to S$ such that any other factor map $\varphi: \Sigma_B \to S$ (from a SFT onto S) must factor through π:

$$
\begin{array}{ccc}
\Sigma_B & \xrightarrow{\;\theta\;} & \Sigma_A \\
\varphi \searrow & & \swarrow \pi \\
& S &
\end{array}
$$

Our main result (Corollary 13) is that a sofic system S has a minimal cover if and only if S is *almost of finite type* (i.e., S can be realized by a factor map which makes identifications on only a proper closed subset of an SFT). This concept (AFT) was introduced in [M] for coding purposes; the theory of modulation codes provides many natural AFT examples. The results herein give an effective method (Corollary 11) of determining whether or not a given sofic system is AFT; this same criterion was found independently by M. Nasu [N2] and W. Krieger (private communication).

Before proceeding to the main result, we review the "future cover" [F1] and note that this cover is minimal (for all sofic systems) among a special class of maps: right closing maps (this follows readily from Fischer's result [F1]). From this minimality result, we recover part of W. Krieger's theorem (Theorem 6) that the "future cover" is a topological conjugacy invariant of the underlying sofic system.

While some of the results of this paper have been discovered independently, one purpose of this paper is to show how they can be obtained from the perspective of minimal covers.

1. Background. We first briefly summarize the relevant background. See [AM, §3] for more details.

Let $\mathscr{A}$ denote $\{1,\ldots,n\}$ with the discrete topology. The full n-shift is the space $\mathscr{A}^Z$ with the product topology. So, a point in the full n-shift is simply a bi-infinite

---

sequence of integers in $\mathscr{A}$. One defines the shift map

$$\sigma: \Sigma_n \to \Sigma_n,$$

$\sigma(x) = y$ where each $y_i = x_{i+1}$. One easily sees that σ is continuous.

By a *subshift* (or symbolic system), we mean the restriction of σ to a closed σ-invariant subset Λ of Σ_n. This means that there is a collection (possibly infinite) C of finite words in the alphabet $\mathscr{A}$ such that (x belongs to Λ) $\Leftrightarrow$ (each finite block of x belongs to C).

Note that since a subshift Λ is defined to be shift invariant, the shift map σ naturally restricts to a map $\sigma: \Lambda \to \Lambda$. Sometimes, we use Λ to refer to either Λ or Λ together with σ/Λ.

Let Λ_1 and Λ_2 be two subshifts with possibly different alphabets. A *factor map* from Λ_1 to Λ_2 is a continuous onto map $\pi: \Lambda_1 \to \Lambda_2$ such that $\pi\sigma = \sigma\pi$.

While the definition of a factor map has an abstract form, it is really a very concrete idea: it is simply a sliding block code. In general, a k-block factor map π is a sliding block code generated by a map

$$\pi^*: \{ k\text{-blocks of } \Lambda_1 \} \to \{ \text{symbols of } \Lambda_2 \},$$

$$\pi((x_i)) = \cdots \pi^*(x_1 \cdots x_k)\pi^*(x_2 \cdots x_{k+1}) \cdots.$$

Factor maps which are 1-1 and onto are called (topological) *conjugacies*. They play a very special role. If a conjugacy from Λ_1 to Λ_2 exists, we say that Λ_1 and Λ_2 are conjugate. The idea is that two conjugate subshifts are essentially the same even if they produce literally different sequences.

Let Λ be a subshift and let k be a positive integer. Let C_k denote the set of all k-blocks of Λ. Define

$$\phi: \Lambda \to (C_k)^z,$$

$$\phi(\cdots x_{-1}x_0x_1 \cdots) = \cdots (x_{-1} \cdots x_{k-2})(x_0 \cdots x_{k-1})(x_1 \cdots x_k) \cdots.$$

Note that the blocks here overlap. The image of ϕ is a subshift, conjugate (*via* ϕ) to Λ. This subshift is called the *higher k-block system* for Λ and is one of many different and convenient ways that we can recode a subshift.

We are interested in subshifts of finite type and sofic systems. A subshift Λ is of *finite* type (SFT) if there is a positive integer k and a collection of k-blocks C such that

$$\Lambda = \{ x \in \mathscr{A}^z : \forall i \; x_{i+1}x_{i+2} \cdots x_{i+k} \in C \}.$$

In other words, Λ is the set of points all of whose k-blocks are prescribed by C.

If $k = 2$, then one constructs an $n \times n$ matrix

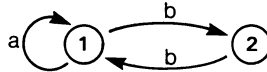$$A_{ij} = \begin{pmatrix} 1 \text{ if } ij \in C \\ 0 \text{ if } ij \notin C \end{pmatrix}.$$

In this case ($k = 2$), the SFT is denoted Σ_A. By a simple recoding (*via* the higher k-block system), every SFT may be described as a Σ_A (with perhaps a much larger set of states). If $ij \in C$, we say that there is an (allowable) transition $i \to j$.

As is standard, one may represent a SFT Σ_A as the set of all bi-infinite walks on a directed graph as follows: The states are the elements of $\mathscr{A}$; one draws an *edge* from i to j if and only if $A_{ij} = 1$. Thus, the point $x = (\cdots x_{-1}x_0x_1x_2 \cdots)$ corresponds to a walk which at time i is at state x_i. For example, if $A = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ then the graph is

One says that a subshift Λ is *sofic* if there is a SFT Σ_A and a factor map π from Σ_A onto Λ. Of course, every SFT is sofic (let π = identity), but sofic systems are much more general.

In the definition of sofic system, by replacing Σ_A by a higher block system, one may assume that π is a 1-block map, or equally well, a 2-block map. From this point of view then, a sofic system is a subshift obtained by labelling the vertices if 1-block (or edges if 2-block) of a directed graph. For example, the edge labelling

generates a sofic system called the even system (b's appear in blocks of even length).

The entropy $h(\Lambda)$ of a subshift Λ is simply the asymptotic growth rate of the number of k-blocks of Λ (as $k \to \infty$). For an irreducible SFT Σ_A, $h(\Sigma_A) = \log(\lambda)$ where λ is the largest eigenvalue of A.

An SFT Σ_A is called *irreducible* if its graph is strongly connected. In the case that Σ_A is not irreducible, it may be studied in terms of its "irreducible components."

2. Right resolving covers. *We assume that all of our sofic systems are transitive with periodic points dense*; every such sofic system is a factor of an *irreducible* SFT (see [CP2]).

DEFINITION. Let $\pi_1: S_1 \to S_1'$ and $\pi_2: S_2 \to S_2'$ be onto factor maps between sofic systems. We say π_1 is *topologically conjugate* to π_2 if there exist topological conjugacies φ and ψ (of the underlying shifts) such that

$$
\begin{array}{ccc}
S_1 & \overset{\psi}{\to} & S_2 \\
\pi_1 \downarrow & & \downarrow \pi_2 \\
S_1' & \overset{\varphi}{\to} & S_2'
\end{array}
$$

commutes.

DEFINITIONS [AM,K]. A factor map $\pi: \Sigma_A \to S$ from a SFT onto a sofic system is *right closing* if whenever x and y are negatively asymptotic points in Σ_A and $\pi(x) = \pi(y)$, then $x = y$. A 1-block factor map π is called *right resolving* if whenever s_1s_2 is a 2-block in S and i is a symbol in Σ_A with $\pi(i) = s_1$, then there is at most one symbol j in Σ_A such that ij is an allowable 2-block in Σ_A and $\pi(ij) = s_1s_2$.

Note. If S is a SFT, then in the above definition "at most one" can be replaced by "exactly one."

Clearly any right resolving map is right closing, but up to a recoding, the converse is true as well. A precise statement of this is as follows:

PROPOSITION 1 (ESSENTIALLY IN [K]). *Every right closing factor map* $\pi\colon \Sigma_A \to S$ *is topologically conjugate to a right resolving factor map* $\pi'\colon \Sigma_{A'} \to S$.

PROOF. By recoding, we may assume that π is a 1-block map. Since π is right closing, a standard compactness argument shows that there is a largest integer M for which there exists a pair of M-blocks, $x_1 \cdots x_M$ and $y_1 \cdots y_M$, such that $x_1 = y_1$, $x_2 \neq y_2$ and $\pi(x_1 \cdots x_M) = \pi(y_1 \cdots y_M)$. We then define $\Sigma_{A'}$ by the partition of M-blocks on Σ_A:

$$x_1 \cdots x_M \sim y_1 \cdots y_M \Leftrightarrow \pi(x_1 \cdots x_M) = \pi(y_1 \cdots y_M) \quad \text{and} \quad x_1 = y_1.$$

We define $\pi'\colon \Sigma_{A'} \to S$ by $\pi'([x_1 \cdots x_M]) = \pi(x_M)$ where [ ] denotes the $\sim$ equivalence class. The maximality condition on M forces π' to be right resolving.

REMARK. One can define the notions of right closing and right resolving for factor maps between arbitrary subshifts, and with a little more work, one can get the result above in this generality.

Similarly, one has the notions of left closing and left resolving.

PROPOSITION 2 [CK]. *Every sofic system S can be realized as the image of a SFT by a 1-1 a.e. right resolving factor map. (Similarly, there is a 1-1 a.e. left resolving factor map which realizes S.)*

We review the construction of Proposition 2 as follows: Let S be a sofic system and let u be a block in S. The *follower set* is defined:

$$\mathcal{F}(u) = \{\text{rays } w \text{ in } S\colon uw \text{ is in } S\}.$$

Let Σ_s^f be the SFT whose *alphabet* is

$$\{(i, \mathcal{F}(u))\colon i \text{ is a symbol in } S \text{ and } u \text{ is a block in } S\}$$

and *transitions* are

$$(i, \mathcal{F}(u)) \to (j, \mathcal{F}(v)) \quad \text{iff} \quad \mathcal{F}(v) = \mathcal{F}(uj).$$

The map $\pi_s^f\colon \Sigma_s^f \to S$ defined by $\pi_s^f((i, \mathcal{F}(u))) \equiv i$ is the 1-1 a.e. right resolving factor map and the pair (Σ_s^f, π_s^f) is called the "future cover."

DEFINITION. Let S be a sofic system. A *magic word* is a block w in S such that whenever u and v are blocks in S and uw and wv are in S, then so is uwv . We leave the following as an exercise (see also [Kr, Lemma 2.7]):

PROPOSITION 3. Σ_s^f *has a unique irreducible component of maximal entropy—namely, the component defined by the symbols*

$$\{(j, \mathcal{F}(w))\colon w = w_1 \cdots w_n \text{ is a magic word and } w_n = j\}.$$

We Let Σ_s^+ denote this component and $\pi_s^+\colon \Sigma_s^+ \to S$ the restriction of the future cover to Σ_s^+. This map is onto.

PROPOSITION 4. *Let $\varphi: \Sigma_B \to S$ be a right closing factor map. Then φ factors through π_s^+, i.e., there exists θ such that*

$$
\begin{array}{ccc}
\Sigma_B & \stackrel{\theta}{\to} & \Sigma_s^+ \\
\varphi \searrow & & \swarrow \pi_s^+ \\
& S &
\end{array}
$$

commutes.

PROOF. By Proposition 1, we may assume that φ is right resolving. Now, in the terminology of [F1], φ represents a "Shannon Graph." In [F1], it is then shown that any "Shannon Graph" can be reduced to a "minimal Shannon Graph" (see the proofs of Theorem 4(iii) and Lemma 3 of [F1]). There, minimal means minimal in terms of the number of vertices, but one easily checks that one has minimality in the sense of the commutative diagram above. The rough idea is that for each symbol i in Σ_B, there is a block w in S such that $\varphi(\mathscr{F}(i)) = \mathscr{F}(w)$. One then defines θ by $\theta(i) = (\varphi(i), \varphi(\mathscr{F}(i)))$ and checks that this is well defined.

PROPOSITION 5. *Suppose $\pi_s^+ = \pi_2 \cdot \pi_1$ where the Range of $\pi_1 =$ Domain of π_2 is a SFT. Then π_1 is 1-1 everywhere.*

PROOF. Let Σ_B denote the range of π_1. We have by hypothesis

$$
\begin{array}{ccc}
\Sigma_s^+ & \stackrel{\pi_1}{\to} & \Sigma_B \\
\pi_s^+ \searrow & & \swarrow \pi_2 \\
& S &
\end{array}
$$

By Proposition 4, we have a diagram

$$
\begin{array}{ccc}
\Sigma_B & \stackrel{\theta}{\to} & \Sigma_s^+ \\
\pi_2 \searrow & & \swarrow \pi_s^+ \\
& S &
\end{array}
$$

Putting these together, we get a diagram

$$
\begin{array}{ccc}
\Sigma_s^+ & \stackrel{\theta \cdot \pi_1}{\to} & \Sigma_s^+ \\
\pi_s^+ \searrow & & \swarrow \pi_s^+ \\
& S &
\end{array}
$$

But since π_s^+ is 1-1 a.e., $\theta \cdot \pi_1 =$ identity a.e. Since $\theta \cdot \pi_1$ is continuous, $\theta \cdot \pi_1 =$ identity. So, π_1 is 1-1. We now recover

THEOREM 6 (KRIEGER [Kr]). *Let S_1 and S_2 be topologically conjugate sofic systems. Then $\pi_{s_1}^+$ and $\pi_{s_2}^+$ are topologically conjugate.*

PROOF. Let $\varphi: S_1 \to S_2$ be the conjugacy. There exists a map $\theta: \Sigma_{s_1}^+ \to \Sigma_{s_2}^+$ by Proposition 4 such that

$$
\begin{array}{ccc}
\Sigma_{s_1}^+ & \overset{\theta}{\to} & \Sigma_{s_2}^+ \\
\pi_{s_1}^+ \downarrow & & \downarrow \pi_{s_2}^+ \\
S_1 & \overset{\varphi}{\to} & S_2
\end{array}
$$

Since $\pi_{s_1}^+ = (\varphi^{-1} \cdot \pi_{s_2}^+) \cdot \theta$, it follows by Proposition 5 that θ is 1-1. Thus, the diagram above is a conjugacy.

Krieger [**Kr**] actually showed that the entire future cover is a conjugacy invariant.

3. AFT systems.

DEFINITION. A sofic system is *almost of finite type* (AFT) if it is the image of an irreducible SFT *via* a factor map which is 1-1 on a nontrivial open set. (Equivalently, the sofic system is realized from the SFT by making identifications on only a proper closed subset.)

The AFT concept was introduced in [**M**] for coding purposes. The following gives an equivalent definition which is easier to work with.

PROPOSITION 7 [**M**]. *Let Σ_A be an irreducible SFT and $\pi: \Sigma_A \to S$ a factor map onto a sofic system S. Then π is 1-1 on an open set if and only if π is right closing, left closing and 1-1 a.e.*

PROPOSITION 8 ([**N1**, THEOREM 6.3]; SEE ALSO [**H**, THEOREM 16.11]). *Let Σ_A and Σ_B be irreducible SFT's. Let $\pi: \Sigma_A \to \Sigma_B$ be right and left closing, then π is constant-to-one.*

DEFINITION. Let $\pi: \Sigma_A \to S$ and $\varphi: \Sigma_B \to S$ be two factor maps from irreducible SFT's Σ_A and Σ_B onto a sofic system S. The *fibered product* of φ and π is the SFT

$$
\Sigma_T = \{(x, y) \in \Sigma_A \times \Sigma_B : \pi(x) = \varphi(y)\}
$$

together with the projection maps $\theta_1: \Sigma_T \to \Sigma_A$, $\theta_2: \Sigma_T \to \Sigma_B$. It is easily verified that Σ_T is indeed a SFT.

The following shows that any AFT sofic system has a minimal cover:

THEOREM 9. *Let S be an AFT sofic system, presented by the factor map $\pi: \Sigma_A \to S$ where π is 1-1 on an open set. Let $\varphi: \Sigma_B \to S$ be any factor map. Assume that Σ_A and Σ_B are both irreducible. Then φ factors through π.*

PROOF. Let Σ_T be the fibered product for π and φ:

$$
\begin{array}{ccc}
 & \Sigma_T & \\
\theta_1 \swarrow & & \searrow \theta_2 \\
\Sigma_A & & \Sigma_B \\
\pi \searrow & & \swarrow \varphi \\
 & S &
\end{array}
$$

We will first show that θ_2 is 1-1 on some irreducible component of Σ_T. To see this, first observe that since π is 1-1 on an open set, it is right closing and left closing (Proposition 7). Thus, so is θ_2. So, θ_2 is finite-to-one (since, by Proposition 1, right closing maps are conjugate to right resolving maps); thus, any irreducible component $\Sigma_{T'}$ (of Σ_T) with maximal entropy (in Σ_T) maps (via θ_2) onto a subshift of Σ_B with maximal entropy in Σ_B. By [**CP1**, Theorem 3.3] this means that $\theta_2|_{\Sigma_{T'}}$ is onto. Now since θ_2 is right and left closing, it follows that $\theta_2|_{\Sigma_{T'}}$ is *constant*-to-one (by Proposition 8). But what *constant* K, can this be? If $K > 1$, then for each point $y \in \Sigma_B$, there exist two points $x, x' \in \Sigma_A$ such that (x, y) and (x', y) belong to $\Sigma_{T'}$ and hence $\pi(x) = \pi(x') = \varphi(y)$. Since y is arbitrary and φ is onto, this then implies that π is at least two-to-1 everywhere. This is contrary to the assumption that π is 1-1 on an open set. Thus, $K = 1$, i.e., θ_2 is invertible. Let $\theta = \theta_1 \circ \theta_2^{-1}$. Then $\varphi = \pi \circ \theta$. So, if we can show that θ is a factor map, we will have shown that φ factors through π. Well, clearly θ is continuous and shift-commuting. To see that θ is onto, first observe that since φ is onto, π maps $\theta(\Sigma_B)$ onto S; since π is finite-to-one, this implies that $\theta(\Sigma_B)$ has maximal entropy in Σ_A. So by [**CP1**, Theorem 3.3], $\theta(\Sigma_B) = \Sigma_A$.

We now derive some consequences: The following shows that an AFT sofic system can be presented as an AFT system in only one way!

COROLLARY 10. *Let S be an AFT sofic system, presented by $\pi: \Sigma_A \to S$ (where π is 1-1 on an open set). Then, π is topologically conjugate to π_s^+.*

PROOF. By the previous theorem, π_s^+ must factor through π:

$$\Sigma_s^+ \xrightarrow{\theta} \Sigma_A$$

$$\pi_s^+ \searrow \qquad \swarrow \pi$$

$$S$$

But then by Proposition 5, θ must be a conjugacy.

The following is a criterion for AFT. (It was also discovered independently by Nasu [**N2**, Theorem 6.1] and W. Krieger.)

COROLLARY 11. *Let S be a sofic system. Then S is AFT if and only if π_s^+ is left closing.*

PROOF. If S is AFT, then by Corollary 10 and Proposition 7, π_s^+ is right and left closing.

Conversely, if π_s^+ is left closing, then since, by construction, it is right closing and 1-1 a.e., S is AFT (by Proposition 7).

LEMMA 12. *Let Σ_A be an irreducible SFT and let $\pi: \Sigma_A \to S$ be a right closing factor map. Suppose $\pi = \pi_2 \cdot \pi_1$ where the (Range of π_1) = (Domain of π_2) is a SFT. Then both π_1 and π_2 are right closing (similarly, for left closing).*

PROOF. Clearly π_1 is right closing. Then by recoding (Proposition 1), we may assume that π_1 is a right *resolving* map from Σ_A onto the (Range of π_1), which is a SFT. Thus, any pair $\{x, y\}$ of negatively asymptotic points in the range of π_1 lifts

(*via* π_1) to a pair of negatively asymptotic points $\{x', y'\}$ in Σ_4. Now, suppose $\pi_2(x) = \pi_2(y)$; we wish to show that $x = y$ (so π_2 will then be right closing). Well,

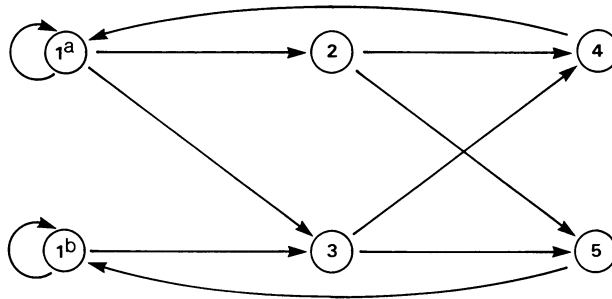$$\pi(x') = \pi_2 \cdot \pi_1(x') = \pi_2(x) = \pi_2(y) = \pi_2 \cdot \pi_1(y') = \pi(y').$$

Since π is right closing, $x' = y'$; thus, $x = y$, as desired.

REMARK. The lemma above is false without the SFT assumption on the Range of π_1. However, the SFT assumption on the domain of π can be relaxed to the assumption that the domain is a sofic system, transitive with periodic points dense.

COROLLARY 13. *A sofic system S is AFT if and only if it has a minimal cover.*

PROOF. *Only if* is Theorem 9. If S has a minimal cover π, then both right and left closing covers of Proposition 2 must factor through π. Hence, by Lemma 12, π must be both right and left closing. It must also be 1-1 a.e. since the right and left closing covers are. Thus, by Proposition 7, S is AFT.

EXAMPLE. We now give a simple example of a sofic system which has no minimal SFT cover—equivalently (by Corollary 13), is *not* AFT. Let S be the sofic system on symbols $1, 2, 3, 4, 5$ obtained by identifying 1^a and 1^b in the graph below:

The reader can easily check that this represents the "future cover" π_s^+ and the covering map is not left closing. So, by Corollary 11, S is not AFT.

Actually, this example was the motivation for this paper; several years ago, one of the authors (B. Kitchens) gave a direct proof that the sofic system S above has no minimal SFT cover. Later one of the other authors (M. Boyle) showed directly (in his Ph.D. thesis), that S is *not* AFT. Now, we know that these two results are equivalent.

We finally raise the following:

QUESTION.[2] For every sofic system S, does there exist a finite collection of SFT 1-1 a.e. covers $\pi_1, \dots, \pi_n$ such that every 1-1 a.e. factor map from a SFT onto S factors through some π_i? Of course, if S is AFT, then one needs only one π_i.

REFERENCES

[AM] R. Adler and B. Marcus, *Topological entropy and equivalence of dynamical systems*, Mem. Amer. Math. Soc., No. 219 (1979).

[CK] I. Csizar and J. Komlos, *On the equivalence of two models of finite-state noiseless channels from the point of view of the output*, Information Theory, Vol. 1, Proc. Colloq. Math. Soc. Janos Bolyai, 1968.

---

[2] S. Williams has answered this question in the negative.

[**CP1**] E. Coven and M. Paul, *Endomorphisms of irreducible subshifts of finite type*, Math. Systems Theory **8** (1974), 167–175.

[**CP2**] _____, *Sofic systems*, Israel J. Math. **20** (1975), 165–177.

[**F1**] R. Fischer, *Graphs and symbolic dynamics*, Information Theory, Vol. 16, Proc. Colloq. Math. Soc. Janos Bolyai, 1975.

[**F2**] _____, *Sofic systems and graphs*, Monatsh. Math. **80** (1975), 179–186.

[**H**] G. A. Hedlund, *Endomorphisms and automorphisms of the shift dynamical system*, Math. Systems Theory **3** (1969), 320–375.

[**K**] B. Kitchens, Ph.D. Thesis, University of North Carolina, Chapel Hill, N.C., 1981.

[**Kr**] W. Krieger, *On sofic systems*. I, Israel J. Math. **48** (1984), 305–330.

[**M**] B. Marcus, *Sofic systems and encoding data*, IEEE—Information Theory **31** (1985), 366–377.

[**N1**] M. Nasu, *Constant-to-one and onto global maps of homomorphisms between strongly connected graphs*, Ergodic Theory Dynamical Systems **3** (1983), 387–414.

[**N2**] _____, *An invariant for bounded-to-one factor maps between sofic subshifts*, Ergodic Theory Dynamical Systems **5** (1985), 89–107.

[**W**] B. Weiss, *Subshifts of finite type and sofic systems*, Monatsh. Math. **77** (1973), 462–474.

MATHEMATICAL SCIENCES RESEARCH INSTITUTE, BERKELEY, CALIFORNIA 94720

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MARYLAND, COLLEGE PARK, MARYLAND (Current address of Mike Boyle)

IBM RESEARCH LABORATORY, YORKTOWN HEIGHTS, NEW YORK 10598 (Current address of Bruce Kitchens)

IBM RESEARCH LABORATORY, SAN JOSE, CALIFORNIA 95193 (Current address of Brian Marcus)