HOMOTOPY HYPERBOLIC 3-MANIFOLDS ARE HYPERBOLIC

David Gabai, G. Robert Meyerhoff, and Nathaniel Thurston

Chapter 0: Introduction

This paper introduces a rigorous computer-assisted procedure for analyzing hyperbolic 3-manifolds. This procedure is used to complete the proof of several long-standing rigidity conjectures in 3-manifold theory as well as to provide a new lower bound for the volume of a closed orientable hyperbolic 3-manifold.

Theorem 0.1: Let N be a closed hyperbolic 3-manifold. Then

i) If $f\colon M \to N$ is a homotopy equivalence, where M is a closed irreducible 3-manifold, then f is homotopic to a homeomorphism.

ii) If $f, g\colon M \to N$ are homotopic homeomorphisms, then f is isotopic to g.

iii) The space of hyperbolic metrics on N is path connected.

Remarks: Under the additional hypothesis that M is hyperbolic, conclusion i) follows from Mostow's rigidity theorem [Mo]. Under the hypothesis that N is Haken (and not necessarily hyperbolic), conclusions i)-ii) follow from Waldhausen [Wa]. Under the hypothesis that N is both Haken and hyperbolic, conclusion iii) follows by combining [Mo] and [Wa]. Because non-Haken manifolds are necessarily orientable we will from now on assume that all manifolds under discussion are orientable.

Theorem 0.1 with the added hypothesis that some closed geodesic $\delta \subset N$ has a *non-coalescable insulator family* was proven by Gabai (see [G]). Thus Theorem 0.1 follows from [G] and the main technical result of this paper which is,

Theorem 0.2: If δ is a shortest geodesic in a closed orientable hyperbolic 3-manifold, then δ has a non-coalescable insulator family.

Remarks: If δ is the core of an embedded hyperbolic tube of radius $\ln(3)/2 = 0.549306...$ then δ has a non-coalescable insulator family by Lemma 5.9 of [G]. (See the Appendix to this paper for a review of insulator theory.) In this paper we establish a second condition, sufficient to guarantee the existence of a non-coalescable insulator family for δ : That *Corona*$(\delta) < 2\pi/3$. (Corona$(\delta) < 2\pi/3$ if tuberadius$(\delta) > \ln(3)/2$). We use the expression "N satisfies the *insulator condition*" when there is a geodesic δ which has a non-coalescable insulator family.

We prove Theorem 0.2 by first showing that all closed hyperbolic 3-manifolds, with seven families of exceptional cases, have embedded hyperbolic tubes of radius $\ln(3)/2$ about their shortest geodesics. (Conjecturally, up to isometry, there are exactly six exceptional manifolds associated to these seven families; see Conjecture 1.31 and Remarks 1.32.) Second, we show that any shortest geodesic δ in six of the seven families has Corona$(\delta) < 2\pi/3$. Finally, we show that the seventh family corresponds to Vol3, the closed hyperbolic 3-manifold with (conjecturally) the third smallest volume, and that the insulator condition holds for Vol3. Each of the three parts of the proof is carried out with the assistance of a rigorous computer program.

Here is a brief description of why Theorem 0.2 might be amenable to computer-assisted proof. If a shortest geodesic δ in a hyperbolic 3-manifold N does not have a $\ln(3)/2$ tube then there is a 2-generator subgroup G of $\pi_1(N) = \Gamma$ which also does not have that property. Specifically, G generated by f and w, with $f \in \Gamma$ a primitive hyperbolic isometry whose fixed axis $\delta_0 \subset \mathbf{H}^3$ projects to δ, and with $w \in \Gamma$ a hyperbolic isometry which takes δ_0 to a nearest translate. Then, after identifying $N = \mathbf{H}^3/\Gamma$ and letting $Z = \mathbf{H}^3/G$, we see that the shortest geodesic in Z (which corresponds to δ) does not have a $\ln(3)/2$ tube. Thus, to understand solid tubes around shortest geodesics in hyperbolic 3-manifolds, we need to understand appropriate 2-generator groups, and this can be done by a parameter space analysis as follows. (Parameter space analyses are naturally amenable to computer proofs.)

The space of relevant 2-generator groups in $\text{Isom}_+(\mathbf{H}^3)$ is naturally parametrized by a subset $\mathcal{P}$ of $\mathbf{C}^3$. Each parameter corresponds to a 2-generator group G with specified generators f and w, and we call such a group a *marked group*. The marked groups of particular interest are those in which G is discrete, torsion free, parabolic free, f corresponds to a shortest geodesic δ, and w corresponds to translation of a particular lift of δ to a nearest translate. We show that if $\text{tuberadius}(\delta) < \ln(3)/2$ in a hyperbolic 3-manifold N, then G must correspond to a parameter lying in one of 7 small regions in $\mathcal{P}$. Roughly speaking, we subdivide $\mathcal{P}$ into a billion small regions and show that all but 7 regions cannot contain such a group G. For example we would know that a region $\mathcal{R}$ contained no such group if we knew that for each point $\rho \in \mathcal{R}$, $\text{length}(f_\rho) > \text{length}(w_\rho)$. (Here $\text{length}(f_\rho)$ (resp. $\text{length}(w_\rho)$) denotes the real translation length of the isometry of $\mathbf{H}^3$ corresponding to the element f (resp. w) in the marked group with parameter ρ.) This inequality would contradict the fact that f corresponds to δ which is a *shortest* geodesic.

Having eliminated from consideration the entire relevant parameter space with the exception of seven small regions, we next "perturb" the computer analysis to eliminate six of the seven regions, that is, all but the region X_0. We do this by showing that if δ is a shortest geodesic in N, and $\text{Corona}(\delta) > 2\pi/3$, then $\pi_1(N)$ has a marked subgroup which lies in X_0. This involves a second computer analysis similar to the first. Here we are interested in discrete, torsion-free, parabolic-free, marked groups $\{G, f, w\}$ where f corresponds to an oriented shortest geodesic δ, and w corresponds to an isometry which minimizes the function $\mathcal{C}(d(\delta_0, h(\delta_0)))$ where $h \in G$, and finally $\mathcal{C}(d(\delta_0, w(\delta_0))) \geq 2\pi/3$. Here δ_0 is a lift of δ, $d(\delta_0, h(\delta_0))$ is the complex distance between the two oriented geodesics $\delta_0, h(\delta_0)$ and $\mathcal{C}$ is a function called the Corona function. Thanks to the first search, we need only analyze a vastly smaller parameter space than $\mathcal{P}$.

Finally, a detailed analysis of the region X_0 shows that X_0 contains a unique group of interest, and that the quotient of $\mathbf{H}^3$ by this group is Vol3. Since Vol3 nontrivially covers no manifold [JR], it follows that if $\text{Corona}(\delta) > 2\pi/3$, then $N = \text{Vol3}$. Through a direct analysis of the geometry of Vol3 we show that Vol3 satisfies the insulator condition.

This paper is organized as follows. In Chapter 1 we describe a space $\mathcal{P}' \subset \mathbf{C}^3$ which naturally parametrizes all relevant marked groups. We explain how a theorem of Meyerhoff as well as elementary hyperbolic geometry considerations imply that we need only consider a compact portion $\mathcal{P}$ of $\mathbf{C}^3$. We explain in detail the plan for proving Theorem 0.2. We will actually be working in the parameter space $\mathcal{W} \supset \exp(\mathcal{P})$. The technical reasons for working in $\mathcal{W}$ rather than in $\mathcal{P}$ are described near the end of Chapter 1. In Chapter 2 we describe and prove the necessary results about the Corona function. In Chapter 3 we prove that the exceptional set X_0 in $\mathbf{C}^3$ contains only Vol3. Also, we prove that if δ is a shortest geodesic in a closed orientable hyperbolic 3-manifold N and Corona(δ) $\geq 2\pi/3$, then $N =$ Vol3. Finally, we show that Vol3 satisfies the insulator condition. In Chapter 4, we prove some applications, one of which is discussed briefly below.

In Chapters 5 through 8 we address the computer-related aspects of the proof. In Chapter 5, the method for describing the decomposition of the parameter space $\mathcal{W}$ into sub-regions is given, and the conditions used to eliminate all but seven of the sub-regions are discussed. Near the end of this chapter, the first part of a detailed example is given. Eliminating a sub-region requires that a certain function is shown to be bounded appropriately over the entire sub-region. This is carried out by using a first-order Taylor series approximation of the function together with a remainder bound. Our computer version of such a Taylor series with remainder bound is called an *AffApprox* and in Chapter 6, the relevant theory is developed. At this point, the detailed example of Chapter 5 can be completed.

As an aside, we note that at the time of this research we believed (based largely on discussions with experts in the field) that there were no available appropriate Taylor series packages. Since carrying out our research, we have discovered that L. Figueiredo and J. Stolfi have independently developed an Affine Arithmetic package which is different in spirit and in the specifics of the implementation than ours, but covers similar ground. One can consult [FS] for an alternate approach to ours.

Finally, in Chapters 7 and 8, round-off error analysis appropriate to our set-up is introduced. Specifically, in Chapter 8, round-off error is incorporated into the *AffApprox* formulas introduced in Chapter 6. The proofs here require an analysis of round-off error for complex numbers, which is carried out in Chapter 7.

We used two rigorous computer programs in our proofs—*verify* and *corona*. These programs are provided at the *Annals* web site. It should be noted that *corona* is a small variation of *verify* and as such only a small number of sections differ from those of *verify*. The proofs of Propositions 1.28, 2.8, and 3.2 amount to having *verify* and *corona* analyze several computer files. These computer files are also available at the *Annals* web site. Details about how to get them and the programs can be found there.

One consequence of our work is

Theorem 4.1: If δ is a shortest geodesic in the closed orientable hyperbolic 3-manifold N, then either

i) Tuberadius(δ) $> \ln(3)/2$, or

ii) $1.0953/2 >$ Tuberadius(δ) $> 1.0591/2$ and length(δ) > 1.059, or

iii) Tuberadius(δ) $= 0.8314\ldots/2$ and $N =$ Vol3.

Combining Theorem 4.1 with [GM2], which implies that if the closed orientable hyperbolic 3-manifold N has a geodesic with a $\ln(3)/2$ tube then the volume of N is greater than $0.16668\ldots$, we obtain

Corollary 4.3: If N is a closed orientable hyperbolic 3-manifold, then the volume of N is greater than $0.16668\ldots$.

Remarks: i) The previous best lower bound for volume was 0.001 by [GM1], which improved the lower bound of 0.0008 of [M2].

ii) Using Theorem 4.1, A. Przeworski has recently extended the lower bound to 0.276796 (see [Pr]).

Given the fundamental use of computers in our proof, we need to discuss issues related to their use. For an introduction to this topic we suggest [La], especially the concluding remarks.

We pose the simple question: why should one have confidence in our proof?

First, the non-computer part of the proof has been analyzed in the traditional way by the authors, referees, individuals, and in seminars.

Second, the computer programs we have written can be checked just as mathematical proofs can be checked, and have been so checked. However, one must be prepared for subtleties that the computational approach introduces. For example, if we wish to have the computer show that the result x of a calculation is less than 2, it is not equivalent to show that x is not greater than or equal to 2. That is because the output of the computation may be "NaN" (not a number) and said output is not "greater than or equal to 2". This may arise, for example, if at some point of the (theoretical) computation one takes the quotient of two numbers, both of which are extremely small. The computer will view both the numerator and the denominator as 0 and hence produce the NaN output. See sections 5, 6, and 7 of [IEEE] for more details.o reconstruct the data set (that is, the decomposition of the parameter space into sub-boxes together with killerwords/conditions), because the hardest part of constructing the data set was the search for killerwords. Presumably, this search could also be done fairly quickly by breaking the parameter space up into small pieces and farming these pieces out to various computers. Also, the locations in the parameter space that caused the most trouble are explicitly described in Proposition 1.28. Knowing these trouble spots and how to deal with them ahead of time, would be a time-saver.

In the unlikely event that "conditionlist" was not available, then the task of reconstructing the data set would be considerably harder. Although, the facts that it *has* been done, that faster and faster computers will be plentifully available, and that Proposition 1.28 saves some work, indicate that the reconstruction process would not be too horrendous. Further, it is also possible that improved proof techniques to the main theorem of this paper will be developed. In fact, Chapter 2 describes a process, *corona*, that has the potential to easily handle the worst of the trouble spots in the parameter space.

Acknowledgements: We thank The Geometry Center and especially Al Marden and David Epstein for the vital and multifaceted roles they played in this work. We also thank the Boston College Physics Department for allowing us to use their suite of computers.

Jeff Weeks and *SnapPea* provided valuable data and ideas. In fact, the data from an undistributed version of *Snappea* encouraged us to pursue a computer-assisted proof of Theorem 0.2.

Bob Riley specially tailored his program *Poincaré* to directly address the needs of our project. His work provided many leads in our search for killerwords. Further, he provided the first proof to show (experimentally) that the six exceptional regions (other than the

Vol3 region) correspond to closed orientable 3-manifolds. The authors are deeply grateful for his help.

The first-named author thanks the NSF for partial support. Some of the first author's preliminary ideas were formulated while visiting David Epstein at the University of Warwick Mathematics Institute. The second-named author thanks the NSF and Boston College for partial support; the USC and Caltech Mathematics Departments for supporting him as a visitor while much of this work was done; and Jeff Weeks, Alan Meyerhoff, and especially Rob Gross for computer assistance. The third-named author thanks the NSF for partial support; and the Geometry Center and the Berkeley Mathematics Department for their support.

Finally, we thank the referees for the magnificent job they did. The first set of referees read our paper thoroughly and made numerous excellent suggestions for improving the exposition. Further, their discussion of issues related to computer-aided proofs crystallized many of these topics in our minds.

The second set of referees also read the paper thoroughly, and we are grateful for their elegant suggestions concerning the exposition. They also checked the programs in great detail, and approached this task with a desire to understand what was really going on behind the scenes. Their ingenious robustness checks raise the confidence level in our proof, and their thought-provoking comments should help us when we attempt to use the computer to help us push across the frontier of our current results.

Chapter 1: Killerwords and the Parameter Space

Notation and Conventions 1.1: A hyperbolic 3-manifold is a Riemannian 3-manifold of constant sectional curvature -1. All hyperbolic 3-manifolds under consideration will be closed and orientable. We will work in the upper-half-space model for hyperbolic 3-space: $\mathbf{H}^3 = \{(x, y, z) : z > 0\}$ with metric $\mathrm{ds_H} = \mathrm{ds_E}/z$. The distance between two points w and v in $\mathbf{H}^3$ will be denoted $\rho(w, v)$.

It is well known that $\mathrm{Isom}_+(\mathbf{H}^3) = \mathrm{PSL}(2,\mathbf{C})$, where an element of PSL(2,**C**) acts as a Möbius transformation on the bounding (extended) complex plane and the extension to upper-half-space is the natural extension (see [Beardon]). If M is a hyperbolic 3-manifold, then $M = \mathbf{H}^3/\Gamma$ where Γ is a discrete, torsion-free subgroup of PSL(2,**C**).

For computational convenience, we will often normalize so that the (positive) z-axis is the axis of an isometry. As such, we set up some special notation. Let $B_{(0;\infty)}$ denote the oriented geodesic $\{(0, 0, z) : 0 < z < \infty\}$, with negative endpoint $(0, 0, 0)$. (An *endpoint* of an axis refers to a limit point of the axis on S^2_∞.) Let $B_{(-1;1)}$ denote the oriented geodesic with negative endpoint $(-1, 0, 0)$ and positive endpoint $(1, 0, 0)$.

When working in a group G generated by f and w and looking at words in f, w, f^{-1}, w^{-1} we will often let F and W denote f^{-1} and w^{-1}, repectively.

Definition 1.2: If f is an isometry, then we define *Relength*$(f) = \inf\{\rho(w, f(w)) \mid w \in \mathbf{H}^3\}$. Thus Relength$(f) = 0$ if and only if f is either a parabolic or elliptic isometry. If Relength$(f) > 0$, then f is hyperbolic and maps a unique geodesic σ in $\mathbf{H}^3$ to itself. In that case σ is oriented (the negative end being the repelling fixed point on S^2_∞) and the isometry f is the composition of a rotation of t (mod 2π) radians along σ (the sign of the angle of rotation is determined by the right-hand rule) followed by a pure translation of

$\mathbf{H}^3$ along σ of $l = \mathrm{Relength}(f)$. We define *length*$(f) = l + it$, and call $A_f = \sigma$ the *axis* of f. A_f is an oriented interval with endpoints in S^2_∞, the orientation being induced from σ.

If the geodesic σ is given a fixed orientation, we define an $l+it$ translation f along σ to be a distance l translation in the positive direction, followed by a rotation of σ by t radians. Of course if $l < 0$, then each point of σ gets moved $-l$ in the negative direction. Also, via the right-hand rule, the orientation determines what is meant by a t-radian rotation. Thus if $l > 0$, then the orientation induced on σ by f (as in the previous paragraph) equals the given orientation. If $l < 0$, then the induced orientation is opposite to the given orientation and f is a $-(l + it)$ translation of $-\sigma$ in the sense of the previous paragraph.

If f is elliptic, then f is a rotation of t radians where $0 \leq t \leq \pi$ about some oriented geodesic, and we define $\mathrm{length}(f) = ti$. If f is parabolic or the identity, we define $\mathrm{length}(f) = 0 + i0$. So, for all isometries we have that Relength = Re(length).

Definition 1.3: If G is a subgroup of $\mathrm{Isom}_+(\mathbf{H}^3)$, then we say that f is a *shortest* element in G if $f \neq \mathrm{id}$ and $\mathrm{Relength}(f) \leq \mathrm{Relength}(g)$ for all $g \in G,\ g \neq \mathrm{id}$.

Definition 1.4: If $\sigma,\ \tau$ are disjoint oriented geodesics in $\mathbf{H}^3$ which do not meet at infinity, then define *distance*$(\sigma, \tau) = \mathrm{length}(w)$ where $w \in \mathrm{Isom}_+(\mathbf{H}^3)$ is the hyperbolic element which translates $\mathbf{H}^3$ along the unique common perpendicular between σ and τ and which takes the oriented geodesic σ to the oriented geodesic τ. The oriented common perpendicular from σ to τ is called the *orthocurve* between σ and τ. The *ortholine* between σ and τ is the complete oriented geodesic in $\mathbf{H}^3$ which contains the orthocurve between σ and τ.

If $\sigma,\ \tau$ intersect at one point in $\mathbf{H}^3$ then there is an elliptic isometry w taking σ to τ fixing $\sigma \cap \tau$. Again, define $\mathrm{distance}(\sigma, \tau) = \mathrm{length}(w)$. In this case, the orthocurve is the point $\sigma \cap \tau$, and the ortholine O from σ to τ is oriented so that $\sigma,\ \tau,\ O$ form a right-handed frame.

If $\sigma,\ \tau$ intersect at infinity, then there is no unique common perpendicular, hence no ortholine, and we define $\mathrm{distance}(\sigma, \tau) = 0 + i0$, or $0 + i\pi$ depending on whether or not σ and τ point in the same direction at their intersection point(s) at infinity.

Define *Redistance* = Re(distance).

As defined, Redistance is non-negative. In Definition 1.8 and in Chapters 2 and 3, it will be useful to have a broader definition. Given an oriented geodesic α in $\mathbf{H}^3$ orthogonal to oriented geodesics β and γ, define $d_\alpha(\beta, \gamma) \in \mathbf{C}$ where a $d_\alpha(\beta, \gamma)$ translation of $\mathbf{H}^3$ along α takes β to γ.

Lemma 1.5:

i) $\mathrm{distance}(\sigma, \tau) = \mathrm{distance}(\tau, \sigma)$

ii) $d_\alpha(-\beta, \gamma) = d_\alpha(\beta, -\gamma) = d_\alpha(\beta, \gamma) + \pi i$

iii) $d_\alpha(\beta, \gamma) = d_{-\alpha}(\gamma, \beta) = -d_\alpha(\gamma, \beta)$. ∎

Definition 1.6: A tube of radius r about a geodesic δ_0 in $\mathbf{H}^3$ is $\{w \in \mathbf{H}^3 \mid \rho(w, v) \leq r$ for some $v \in \delta_0\}$. If δ is a simple closed geodesic in the hyperbolic 3-manifold N and if $\{\delta_i\}$ is the set of pre-images of δ in $\mathbf{H}^3$, then define *tuberadius*$(\delta) = \frac{1}{2} \min \mathrm{Redistance}(\delta_i, \delta_j) \mid i \neq j\}$. If $r = \mathrm{tuberadius}(\delta)$, then define a *maximal tube* about δ to be the image of a tube of radius r about δ_0. Note that $\mathrm{tuberadius}(\delta) = \sup\{r \mid$ there exists an embedded tubular neighborhood of radius r about $\delta\}$.

Lemma 1.7: Let δ be a closed geodesic in the hyperbolic 3-manifold N and $\{\delta_i\}_{i\geq 0}$ be the set of its distinct lifts to $\mathbf{H}^3$, then tuberadius$(\delta) = (\frac{1}{2})\min\{\text{Redistance}(\delta_0, \delta_i) \mid i \neq 0\}$. ∎

Definition 1.8: Our desire to understand tuberadii about closed geodesics, and especially about a simple closed geodesic δ, leads us to investigate certain 2-generator subgroups $G =< f, w >$ of $\text{Isom}_+(\mathbf{H}^3)$ with the generator f corresponding to a primitive isometry fixing δ_0 and the generator w corresponding to an element taking δ_0 to its nearest covering translate. We investigate these 2- generator groups by using certain subsets of $\mathbf{C}^3$ as parameter spaces.

A *marked (2-generator) group* is a triple $\{G, f, w\}$ consisting of a 2- generator subgroup G of $\text{Isom}_+(\mathbf{H}^3)$ and an ordered pair of isometries f, w of $\mathbf{H}^3$ which generate G such that Relength$(f) > 0$ and if A_f is the axis of f, then $w(A_f) \cap A_f = \emptyset$ (here, intersection is taken in $\mathbf{H}^3 \cup S^2_\infty$). Two marked groups $\{G_1, f_1, w_1\}$ and $\{G_2, f_2, w_2\}$ are *conjugate* if G_1 and G_2 are conjugate via an element of $\text{Isom}_+(\mathbf{H}^3)$ and this conjugating element takes f_1 to f_2 and w_1 to w_2. Within any conjugacy class of marked groups is a unique normalized element $\{G, f, w\}$ where f is a positive translation along the (oriented) geodesic $B_{(0;\infty)}$, and the orthocurve from $w^{-1}(B_{(0;\infty)})$ to $B_{(0;\infty)}$ lies on $B_{(-1;1)}$ on the negative side of $B_{(-1;1)} \cap B_{(0;\infty)}$. To minimize notation, we will frequently equate a conjugacy class with its normal representative.

Given $(L, D, R) = (l + it, d + ib, r + ia) \in \mathbf{C}^3$ with $l > 0,\ d > 0$, one can associate a group G generated by elements f and w as follows. Define f to be an $l+it$ translation along $B_{(0;\infty)}$ and w to be a $d+ib$ translation along $B_{(-1;1)}$ followed by an $r+ia$ translation along $B_{(0;\infty)}$ (here, r can be negative, in which case this is equivalent to $-r-ia$ translation along $-B_{(0;\infty)}$). Conversely if $\{G, f, w\}$ is a normalized marked group then f is an L translation of $B_{(0;\infty)}$ and w is a D translation of $B_{(-1;1)}$ followed by an R translation of $B_{(0;\infty)}$. Thus $\mathcal{P}' = \{(l + it, d + ib, r + ia) \in \mathbf{C}^3 |\ l > 0, d > 0\}$ parametrizes the set of conjugacy classes of marked groups. In particular, the parametrization is surjective and locally one-to-one.

We are primarily interested in the set $\mathcal{T}' \subset \mathcal{P}'$ which parametrizes all conjugacy classes of marked groups $\{G, f, w\}$ for which f is a shortest element of G which (positively) translates $B_{(0;\infty)}$ and $w \in G$ takes $B_{(0;\infty)}$ to a *nearest* translate $w(B_{(0;\infty)})$ such that $-\text{Relength}(f)/2 < d_{B_{(0;\infty)}}($ (ortholine from $w^{-1}(B_{(0;\infty)})$ to $B_{(0;\infty)}$), (ortholine from $B_{(0;\infty)}$ to $w(B_{(0;\infty)}))) \leq \text{Relength}(f)/2$. See Figure 1.1.

Figure 1.1:

Remark 1.9: $\mathcal{T}'$ consists of those parameters corresponding to marked groups $\{G, f, w\}$ such that l is the real length of a shortest element of G, d is the real distance between $B_{(0;\infty)}$ and a nearest translate, and $-l/2 < r \leq l/2$. In what follows, it is essential to remember that an element α of $\mathcal{P}'$ corresponds not only to a group G, but to a marked group. To further establish the point, we note that the parameter l is an invariant of G alone (that is, l is the shortest real length of an element of G), while the parameter d is determined by G and f (that is, the notion of "nearest" used to define w in the definition of $\mathcal{T}'$ requires a choice of f).

As mentioned in the Introduction to this paper, we are only interested in the subset of $\mathcal{T}'$ corresponding to parameters α with $d \leq \ln(3)$. The following two propositions imply this subset of $\mathcal{T}'$ lives in a compact subset of $\mathcal{P}'$.

Proposition 1.10: All closed geodesics of length less than 0.0979 in all hyperbolic 3-manifolds have (embedded) tubes of radius $\ln(3)/2$.

Proof: In [M1] it is proven that a closed geodesic of length $x+iy$ has a tube (embedded) of radius $r(x+iy)$ satisfying

$$\sinh^2(r(x+iy)) = \max_{n\in\mathbf{Z}_+} \frac{1}{2}(\frac{\sqrt{1-2k(x,y,n)}}{k(x,y,n)} - 1) \text{ where } k(x,y,n) = \cosh(nx) - \cos(ny),$$

where we restrict to $x+iy$ values which produce positive radii $r(x+iy)$ by means of this formula. It is easy to compute that for a given $x+iy$ we need to have n for which $0 < k(x,y,n) < \sqrt{2}-1$ to produce a positive radius tube by this method.

The function $\frac{1}{2}(\frac{\sqrt{1-2k}}{k} - 1)$ is decreasing on the interval $(0, -1+\sqrt{2})$. It is easy to solve for the k value that produces radius $r = \ln(3)/2$ and it is just over 0.3397. Thus, by restricting to k values in the interval $(0, 0.3397)$ we guarantee radii r greater than $\ln(3)/2$.

Thus, to complete the proof of this proposition, we need to show that when a closed geodesic has real length x less than 0.0979, there exists a positive integer n for which $k(x,y,n)$ is less than 0.3397 for all angles y. Because cosh is an increasing function, we can restrict our analysis to $x = 0.0979$. Thus, we need only show that given any angle y, we can find a positive integer n such that $\cosh(n0.0979) - \cos(ny) < 0.3397$. When $n > 8$ we can compute that $\cosh(n0.0979) - \cos(ny) > 0.3397$, and we therefore restrict to positive integers $n \leq 8$.

We now consider angles y. Because cos is an even function, we need only consider $y \in [0,\pi]$. Finally, we complete the proof by covering $[0,\pi]$ by 11 overlapping closed sub-intervals σ_i each of which has an associated positive integer n_i for which $\cosh(n_i 0.0979) - \cos(n_i y) < 0.3397$ is true for all $y \in \sigma_i$:

$$\sigma_0 = [0.000, 0.843] \quad n_0 = 1 \qquad \sigma_5 = [1.733, 1.858] \quad n_5 = 7$$

$$\sigma_1 = [0.835, 0.960] \quad n_1 = 7 \qquad \sigma_6 = [1.832, 2.357] \quad n_6 = 3$$

$$\sigma_2 = [0.951, 1.143] \quad n_2 = 6 \qquad \sigma_7 = [2.334, 2.3792] \quad n_7 = 8$$

$$\sigma_3 = [1.123, 1.391] \quad n_3 = 5 \qquad \sigma_8 = [2.3789, 2.647] \quad n_8 = 5$$

$$\sigma_4 = [1.386, 1.755] \quad n_4 = 4 \qquad \sigma_9 = [2.630, 2.755] \quad n_9 = 7$$

$$\sigma_{10} = [2.730, \pi] \quad n_{10} = 2$$

∎

Proposition 1.11: A *shortest* geodesic δ in a closed hyperbolic 3-manifold has
i) tuberadius$(\delta) \geq l/4$ where $l = $ Relength(δ), and
ii) tuberadius$(\delta) \geq \ln(3)/2$ if Relength$(\delta) \geq 1.289785$.

Proof: Part i is a consequence of the following well-known argument: Uniformly expand a tube around a shortest geodesic in the hyperbolic 3-manifold. If the expanding tube hits itself before a radius of $l/4$ then we will construct a loop of length less than l, which produces a contradiction to being shortest. Drop the two obvious perpendiculars from the hitting point down to the core geodesic. Consider the following loop—down one perpendicular, follow the shorter direction on the core geodesic, up the other perpendicular. Because a lift of this loop to $\mathbf{H}^3$ is not closed, this loop is homotopically nontrivial. By construction it has length less than $l/4 + l/2 + l/4 = l$.

To prove part ii, we improve on this loop. Replace the first half of the journey by the hypotenuse of the right triangle formed by the first perpendicular and the first half of the shorter arc along the core geodesic. Replace the second half of the journey by the hypotenuse of the right triangle formed by the second perpendicular and the second half of the shorter arc along the core geodesic. Using the hyperbolic Pythagorean Theorem (see [F]) $\cosh c = (\cosh a)(\cosh b)$ with $a = \ln(3)/2$ and $b = l/4$ we get that the length of the constructed loop is $2\text{Arccosh}(\cosh(\ln(3)/2)\cosh(l/4))$ and this is less than l when $l > 1.289784...$, by a calculation (which follows) and the fact that $2\text{Arccosh}(\cosh(\ln(3)/2)(\cosh(l/4)) - l$ is a decreasing function of l.

We solve explicitly for the value of l at which

$$2\text{Arccosh}(\cosh(\ln(3)/2)(\cosh(l/4)) - l = 0.$$

Noting that $\cosh(\ln(3)/2) = \frac{2}{\sqrt{3}}$ we get $\frac{2}{\sqrt{3}}(\cosh(l/4))) = \cosh(l/2)$. Using a half-angle formula for $\cosh(l/2)$ we get $\frac{2}{\sqrt{3}}(\cosh(l/4))) = 2\cosh^2(l/4) - 1$. Setting $x = \cosh(l/4)$ we get the quadratic $\frac{2}{\sqrt{3}}x = 2x^2 - 1$. Solving and substituting, we get $l = 1.289784...$ ∎

Definition 1.12: Let $\mathcal{P} \subset \mathcal{P}'$ be the set of those parameters $\alpha = (l + it, d + ib, r + ia)$ such that

a) $0.0978 \leq l \leq 1.289785$
b) $l/2 \leq d \leq \ln(3)$
c) $0 \leq r \leq l/2$
d) $-\pi \leq t \leq 0$
e) $-\pi \leq b \leq \pi$
f) $-\pi \leq a \leq \pi$

Define $\mathcal{T} = \mathcal{T}' \cap \mathcal{P}$.

The point of the definition of $\mathcal{P}$ and $\mathcal{T}$ is as follows. We want to analyze by computer the relationship between lengths of shortest geodesics and their tuberadii in hyperbolic 3-manifolds. We were naturally led to the parameter space $\mathcal{P}'$ and its subset $\mathcal{T}'$. But $\mathcal{P}'$ is problematic from the computational viewpoint because it is non-compact. We wish to replace $\mathcal{P}'$ by $\mathcal{P}$ which is compact, and $\mathcal{T}'$ by $\mathcal{T}$ in our computer analysis. This is carried out in Lemma 1.13. Note that we worked to make $\mathcal{P}$ as small as reasonable to save computation time; for example, the t and r restrictions above cut down the parameter space by a factor of 4 over the obvious t and r restrictions.

Lemma 1.13: If $\alpha = (l + it, d + ib, r + ia) \in \mathcal{T}'$ has $d \leq \ln(3)$ and corresponds to a 2-generator group $\{G_\alpha, f_\alpha, w_\alpha\}$, then there exists a parameter $\beta = (l' + it', d' + ib', r' + ia') \in \mathcal{T}$ with associated group $\{G_\beta, f_\beta, w_\beta\}$ such that G_β is conjugate (in $\text{Isom}(\mathbf{H}^3)$) to G_α.

Proof: We note that $d < l/2$ is eliminated from consideration by Proposition 1.11i, and the definition of $\mathcal{T}'$. If $d \leq \ln(3)$, then $0.0978 \leq l \leq 1.289785$ by Propositions 1.10 and

1.11ii. If for the marked group $\{G, f, w\}$ we have $-l/2 < r < 0$, then the marked group $\{G, f, w^{-1}\}$ is conjugate to an element of $\mathcal{T}$ whose new r-parameter is $-r$. Thus we can assume that conditions a, b, c from Definition 1.12 hold for the relevant $\{G, f, w\}$. Further, conditions e and f hold.

This leaves condition 1.12d. Conjugating G by a reflection in the geodesic plane spanned by $B_{(0;\infty)}$ and $B_{(-1;1)}$ changes the t-parameter to $-t$ (mod 2π) but leaves the r and d parameters unchanged. The effect on b and a is irrelevant. ∎

By [G; Lemma 5.9] (or see Example A.3 in the Appendix) a closed orientable hyperbolic 3-manifold N satisfies the insulator condition provided that tuberadius(δ) $> \ln(3)/2$ for some closed geodesic $\delta \subset N$. Thus we are led to consider

Problem 1.14: List all closed orientable hyperbolic 3-manifolds N possessing a shortest geodesic δ such that tuberadius(δ) $\leq \ln(3)/2$.

Remarks 1.15: i) Using a J. Weeks-modified version of *SnapPea* [W1] it was known experimentally that any shortest geodesic in Vol3 has a 0.415... tube (see [G]). Conjecturally, up to isometry, there are a total of six manifolds in the list answering Problem 1.14 (see Conjecture 1.31 and Remarks 1.32i and 1.32iii, and Conjecture 4.2).

ii) If a shortest geodesic δ in N satisfies tuberadius(δ) $\leq \ln(3)/2$, then N gives rise to an element $\alpha \in \mathcal{T}$. (In fact N may give rise to finitely many different elements of $\mathcal{T}$.) Thus we need to investigate

Problem 1.16: Find all parameters $\alpha = (l + it, d + ib, r + ia) \in \mathcal{T}$.

Remark 1.17: In the next couple of paragraphs, we will describe our method of (partially) answering Problem 1.16. But before starting this description, we mention a technical point: starting with Definition 1.22, we will realize major advantages by working in the space $\mathcal{W} \supset \exp(\mathcal{P})$, and our results will ultimately be described in terms of $\mathcal{W}$. But for now, for simplicity, we will describe the results in terms of the unexponentiated space $\mathcal{P}$.

We will partition $\mathcal{P}$ into about one billion regions $\{\mathcal{P}_i\}$ and show that $\mathcal{T}$ is disjoint from all but seven small such regions. Suppose that $\mathcal{P}_i$ is a region of this partition and $\alpha \in \mathcal{P}_i$. Let h be a word in the letters f, w and their inverses. Associated to the parameter $\alpha = (l_\alpha + it_\alpha, d_\alpha + ib_\alpha, r_\alpha + ia_\alpha)$ there are the group elements f_α, w_α and hence h_α. If h_α is not the identity then we ask

a) Is Relength(h_α) $<$ Relength(f_α) $= l_\alpha$?
b) Is Redistance($h_\alpha(B_{(0;\infty)}), B_{(0;\infty)}$) $<$ Redistance($w_\alpha(B_{(0;\infty)}), B_{(0;\infty)}$) $= d_\alpha$?

If either a) or b) is true, then $\alpha \notin \mathcal{T}$.

Now let $\beta \in \mathcal{P}_i$, with f_β, w_β, and h_β the associated hyperbolic isometries. If say a) is true for α then so is the statement Relength(h_β) $<$ Relength(f_β) $= l_\beta$ for β sufficiently close to α. Thus we can show that $\mathcal{T} \cap \mathcal{P}_i = \emptyset$ if we can find an α for which say statement a) is true, and then use first-order Taylor approximation (with error/remainder term) to show that the corresponding statement holds for all $\beta \in \mathcal{P}_i$ while continuing to avoid the prohibition that h_β is not the identity.

Definition 1.18: A word h in w, f, w^{-1}, f^{-1} for which statement a) (resp. b)) in Remark 1.17 holds for each $\beta \in \mathcal{P}_i$ and h_β is not a power of f_β for each $\beta \in \mathcal{P}_i$ is called a *killerword* for $\mathcal{P}_i$ with respect to contradiction a) (resp. b)).

Summary 1.19: With seven exceptions, to each of the approximately one billion regions partitioning $\mathcal{P}$, we will associate a killerword and a contradiction.

Remark 1.20: Computers are well suited for partitioning a region such as $\mathcal{P}$ into many sub-regions $\{\mathcal{P}_i\}$, and finding a killerword h_i which eliminates all $\alpha_i \in \mathcal{P}_i$ due to contradiction C_i. Depending on the contradiction, we find computable expressions for approximations of the values of $\text{Relength}(h_\beta)$ or $\text{Redistance}(h_\beta(B_{(0;\infty)}), B_{(0;\infty)})$ and thus use the computer to eliminate all of $\mathcal{P}_i$.

There are a number of difficulties in executing this procedure. First, a uniform mesh of the partition would yield far too many sub-regions to be handled by computer. In fact with 6 real parameters, refining a given mesh by a factor of 10 would change the partition size by a factor of 10^6. Our method for refining the parameter space and the way the computer keeps track of the refinements are discussed briefly in Remark 1.26ii and in more detail in Chapter 5.

A second difficulty is finding the killerwords. In practice, most of the parameter space is eliminated by killerwords of length less than 7, but a number of spots use killerwords of length 10 and a few regions use killerwords of length 44. A brute force enumeration and testing of the various words would take far too long. Note that there are 78,732 words of length 10 and 4×3^{29} words of length 30. An outline of the method we used for finding the partition of $\mathcal{P}$ and the associated killerwords is given in Remark 1.34. However, the proofs in this paper do not require the method for finding the partition of $\mathcal{P}$ and the associated killerwords, they merely require having the partition and the associated killerwords.

Finally, there is the issue of rigor. The main difficulty in making the plan work rigorously is that, in computer calculations, we need to account for the difference between the result of an operation and its approximate computed value. The accumulated errors can become significant when performing many operations on small regions, such as when using a 43-letter word on a box of side length 2^{-22}. A substantial portion of this paper is devoted to addressing the issue of rigor. See also Preview 1.35.

Remark 1.21: To analyze Relength and Redistance as in Remark 1.17, we would be led to work with the Arccosh function, because,

$$\text{length}(f) = 2\text{Arccosh}(\text{trace}(A)/2)$$

where $A \in \text{SL}(2, \mathbf{C})$ represents the isometry f. (As we do not need this formula for $\text{length}(f)$ we will neither prove the formula, nor explain technical details about it.) This would be problematic from the view-point of error analysis—we do not want to deal with transcendental functions such as Arccosh.

This problem can be slickly avoided by exponentiating the preliminary parameter space $\mathcal{P}$ to get the parameter space $\mathcal{W}$ (the definition of $\mathcal{W}$ is given in Definition 1.22). Lemmas 1.24 and 1.25 then demonstrate that while working in $\mathcal{W}$ one need only understand the basic arithmetic operations $+, -, \times, /, \sqrt{}$. The machine implementation of these basic operations is governed by the IEEE standard IEEE-754 (see [IEEE]).

To expand a bit on the problematic nature of transcendental functions, we note that our computer version of Taylor series approximations (see Preview 1.35 and Chapters 6 and 8) is designed to work for functions built up out of the basic arithmetic operations. It would be a nightmare to include functions such as the Arccosh.

Definition 1.22: Let

$$\mathcal{W} = \{(x_0, x_1, x_2, x_3, x_4, x_5) : |x_i| \leq 4 \times 2^{(5-i)/6} \text{ for } i = 0, 1, 2, 3, 4, 5\}$$

$$\supset \exp(\mathcal{P}) = \{(x_0, x_1, x_2, x_3, x_4, x_5) \mid \\ x_0 + ix_3 = \exp(e), x_1 + ix_4 = \exp(f), x_2 + ix_5 = \exp(g) \text{where} (e, f, g) \in \mathcal{P}\}$$

and let

$$\mathcal{S} = \exp(\mathcal{T}).$$

As we are taking exp of the various complex co-ordinates, it is notationally convenient to replace our complex parameters $L = l + it,\ D = d + ib,\ R = r + ia$ by exponentiated versions. That is, let

$$L' = \exp(L) = \exp(l + it),\ D' = \exp(D) = \exp(d + ib),\ R' = \exp(R) = \exp(r + ia).$$

Remarks 1.23: i) We work with $\mathcal{W}$ instead of $\exp(\mathcal{P})$ because we want our initial region to be a (6-dimensional) box that is easily sub-divided. This has the side-effect that certain sub-boxes $\mathcal{W}_i$ of $\mathcal{W}$ will be eliminated because they are outside of $\exp(\mathcal{P})$ rather than by the analogues of conditions a) and b) in Remark 1.17. The entire collection of conditions is given in Chapter 5.

ii) The presence of the factor $2^{(5-i)/6}$ in the definition of $\mathcal{W}$ is explained in Construction 5.3. Briefly, the main reason for doing it is to make the Taylor Series approximations efficient, hence fast.

iii) We chose the co-ordinates of $\mathcal{W}$ so that $L' = x_0 + ix_3,\ D' = x_1 + ix_4,\ R' = x_2 + ix_5$ to gain a mild computer advantage.

Lemma 1.24: If $(L', D', R') \in \mathcal{W}$ and $\{G, f, w\}$ is the associated normalized marked group, then f and w have matrix representatives
a)

$$f = \begin{pmatrix} \sqrt{L'} & 0 \\ 0 & 1/\sqrt{L'} \end{pmatrix}$$

b)

$$w = \begin{pmatrix} \sqrt{R'} * ch & \sqrt{R'} * sh \\ sh/\sqrt{R'} & ch/\sqrt{R'} \end{pmatrix}$$

where $ch = (\sqrt{D'} + 1/\sqrt{D'})/2$ and $sh = (\sqrt{D'} - 1/\sqrt{D'})/2$

Proof: a) In our set-up the (oriented) axis of f is $B_{(0;\infty)}$. As such, f corresponds to a diagonal matrix, with diagonal entries p and p^{-1}, with $|p| > 1$. The action of f on the bounding complex plane is simply multiplication by p^2. Extending this action to upper-half-space in the natural way rotates the z-axis by angle $\arg(p^2)$ and sends

$(0,0,1)$ to $(0,0,|p|^2)$. Thus, $\mathrm{Im}(\mathrm{length}(f)) = \arg(p^2) = \mathrm{Im}(\ln(p^2))$ and, using the hyperbolic metric, $\mathrm{Re}(\mathrm{length}(f)) = \ln(|p|^2) = \mathrm{Re}(\ln(p^2))$. That is, $\mathrm{length}(f) = \ln(p^2)$ and $p = \pm\exp(\mathrm{length}(f)/2) = \pm\sqrt{\exp(\mathrm{length}(f))} = \pm\sqrt{\exp(L)} = \pm\sqrt{L'}$. Now, we take the positive square root (taking the negative square root produces the other lift from PSL(2,**C**) to SL(2,**C**)).

b) $w = \beta \circ \alpha$ where β is translation of distance R along $B_{(0;\infty)}$ and α is translation of distance D along $B_{(-1;1)}$. Thus, a matrix representative of β is

$$\begin{pmatrix} \sqrt{R'} & 0 \\ 0 & 1/\sqrt{R'} \end{pmatrix}$$

and a matrix representative of α can be computed to be

$$\begin{pmatrix} \cosh(D/2) & \sinh(D/2) \\ \sinh(D/2) & \cosh(D/2) \end{pmatrix}.$$

But $\cosh(D/2) = (\exp(D/2) + \exp(-D/2))/2 = (\sqrt{D'} + 1/\sqrt{D'})/2 = ch$ and similarly for sh. Thus,

$$\alpha = \begin{pmatrix} ch & sh \\ sh & ch \end{pmatrix}$$

and b) follows by matrix multiplication. ∎

Lemma 1.25: If $h \in \mathrm{Isom}_+(\mathbf{H}^3)$ is represented by the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbf{C}),$$

then

a) $\exp(\mathrm{Relength}(h)) = |\mathrm{trace}(A)/2 \pm \sqrt{(\mathrm{trace}(A)/2)^2 - 1}|^2$

b) $\exp(\mathrm{Redistance}(h(B_{(0;\infty)}), B_{(0;\infty)})) = |\mathrm{orthotrace}(A) \pm \sqrt{(\mathrm{orthotrace}(A))^2 - 1}|$
where $\mathrm{orthotrace}(A) = ad + bc$.

In both cases, the $+, -$ produce reciprocal values for the right-hand side, and we take the one producing the larger value, unless the value is 1, in which case there is no need to choose.

Proof: a) If A is elliptic or parabolic, the proof is straightforward (the trace of a parabolic is ± 2 while the trace of an elliptic is a real number between 2 and -2).

We assume A is hyperbolic. Because trace is a conjugacy invariant, we can assume the oriented axis of A is $B_{(0;\infty)}$. Thus A is a diagonal matrix with p and p^{-1} along the diagonal with $|p| > 1$, and, as in the proof of Lemma 1.24, we see that $\exp(\mathrm{length}(h)) = p^2$. Of course, $\mathrm{trace}(A) = p + p^{-1}$, and it is easy enough to solve for p. Specifically, $p = \mathrm{trace}(A)/2 \pm \sqrt{(\mathrm{trace}(A)/2)^2 - 1}$.
Thus, $\exp(\mathrm{Relength}(h)) = |\exp(\mathrm{length}(h))| = |p|^2 = |(\mathrm{trace}(A)/2) \pm \sqrt{(\mathrm{trace}(A)/2)^2 - 1}|^2$.

b) If $B_{(0;\infty)}$ and $h(B_{(0;\infty)})$ intersect at infinity, then the proof is straightforward. For example, if h fixes the point $(0,0,0)$ at infinity, then $c = 0$, $ad = 1$ and the formula holds. Similarly for the other cases in which $B_{(0;\infty)}$ and $h(B_{(0;\infty)})$ intersect at infinity.

We assume $B_{(0;\infty)}$ and $h(B_{(0;\infty)})$ do not intersect at infinity. We will compute the length of k, the square of the transformation taking $B_{(0;\infty)}$ to $h(B_{(0;\infty)})$ along their ortholine. Let τ be 180-degree rotation about $B_{(0;\infty)}$, then $(h \circ \tau \circ h^{-1})$ is 180-degree rotation about $h(B_{(0;\infty)})$, and we have that $k = (h \circ \tau \circ h^{-1}) \circ \tau$. Now, τ and h are represented by the matrices

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbf{C}),$$

hence, $k = (h \circ \tau \circ h^{-1}) \circ \tau$ can be computed to have matrix representation

$$\begin{pmatrix} ad + bc & 2ab \\ 2cd & ad + bc \end{pmatrix}.$$

Thus, $\exp(\text{Redistance}(h(B_{(0;\infty)}), B_{(0;\infty)})) = \exp(\text{Relength}(k)/2) = \sqrt{|\exp(\text{length}(k)|} = |(\text{trace}(k)/2) \pm \sqrt{(\text{trace}(k)/2)^2 - 1}| = |(ad + bc) \pm \sqrt{(ad + bc)^2 - 1}|$. ■

Remarks 1.26: i) It follows from Lemma 1.25 that if h is a word in f, w, f^{-1}, w^{-1}, then for any parameter value $\alpha \in \mathcal{W}$, $\exp(\text{Relength}(h_\alpha))$, and $\exp(\text{Redistance}(h_\alpha(B_{(0;\infty)}), B_{(0;\infty)}))$ can be computed using only the operations $+, -, \times, /, \sqrt{}$.

ii) During the course of the computer work needed to prove the main theorems, the parameter space $\mathcal{W}$ was decomposed into sub-boxes by computer via a recursive sub-division process: Given a sub-box that is being analyzed, either it can be *killed directly* (that is, eliminated by a killerword and associated condition as described in Remark 1.17, or for the trivial reason described in Remark 1.23i), or it cannot. If it cannot be killed directly, it is sub-divided in half by a hyper-plane $\{x_i = c\}$ (where i runs through the various co-ordinate dimensions cyclically) and the two pieces are analyzed separately. And so on.

As such, a sub-box of $\mathcal{W}$ can be described by a sequence of 0's and 1's where 0 means "take the lesser x_i values" and 1 means "take the greater x_i values." Remarkably, for the decomposition of $\mathcal{W}$ into sub-boxes, all the sub-box descriptions could be neatly encoded into one tree (although in practice we found it preferable to use several trees to describe the entire decomposition). This is described in Chapter 5.

iii) In the following proposition, seven exceptional boxes are described as sequences of 0's and 1's. Four of the boxes—X_0, X_4, X_5, X_6—are each the union of two abutting sub-boxes, $X_0 = X_{0a} \cup X_{0b}$ and so on. It is a pleasant exercise to work through the fact that they abut. It should be noted that had the set-up for $\mathcal{W}$ been different, more boxes (or perhaps less) might have been needed to construct the seven exceptional regions.

It is also a pleasant exercise to calculate by hand the co-ordinate ranges of the various sub-boxes. For example, the range of the last co-ordinate (i.e., x_5) of the sub-box

X_{6a} = 111000000001000111 111111110101001111 011111010111111111
110001001011000111 0

is found by taking the 6th entry, the 12th entry, the 18th entry, and so on. These entries are 011111111111. The first entry (0) means take the lesser x_5 values, and produces the interval $[-4, 0]$. The second entry (1) means take the greater x_5 values, and produces the interval $[-2, 0]$. The third entry (1) produces $[-1, 0]$. Continuing, we see that X_{6a} has $-2^{-9} \leq x_5 = \mathrm{Im}(R') \leq 0$. The other co-ordinates can be computed in the same fashion, although they must at the end be multiplied by the factor $2^{(5-i)/6}$ (see the definition of the initial box $\mathcal{W}$). The range of co- ordinate values for each sub-box $X_0, X_1, \ldots, X_6$ is given in Table 1.1 (a limited number of significant digits is given), and then a range of co-ordinates for boxes $\mathcal{R}_i \supset \exp^{-1}(X_i)$ is given (see Remarks 1.30i and 1.30ii) in Table 1.2. Finally, two *quasi-relators* are given in Proposition 1.28 for each sub-box $X_0, X_1, \ldots, X_6$ (see the next definition).

Definition 1.27: A *quasi-relator* in a sub-box X of $\mathcal{W}$ is a word in $f, w, F = f^{-1}, W = w^{-1}$ that is close to the identity throughout X and experimentally appears to be converging to the identity at some point in X. In particular, a quasi-relator rigorously has Relength less than that of f at all points in X.

Proposition 1.28: Within the parameter space $\mathcal{W}$ but outside the seven exceptional boxes there are no parameter points corresponding to marked groups $\{G, f, w\}$ where G is torsion free and parabolic free and where f corresponds to a shortest geodesic δ of tuberadius $\leq \ln(3)/2$. Specifically, $\mathcal{S} \cap (\mathcal{W} - \bigcup_{n=0,\ldots,6} X_n) = \emptyset$ where the X_n are the exceptional boxes

$X_{0a} =$ 001000110111110001 101001010101011001 011011010111101101
100001101101000111 010001110101100101 1101110111110100

$X_{0b} =$ 001001110110110000 101000010100011000 011010010110101100
100000101100000110 010000110100100100 1101100111100100

X_0 quasi-relators:
$r_1 = fwFwwFwfww$
$r_2 = FwfwfWfwfw$

$X_1 =$ 001000110001110110 011101000110111110 100010110000100011
101101001101001000 110101011000000100 000

X_1 quasi-relators:
$r_1 = FFwFWFWfWFWFwFFww$
$r_2 = FFwwFwfwfWfwfwFww$

$X_2 =$ 001000110101010010 101010110001100101 110111100001101010
111100100000010001 111100

X_2 quasi-relators:

$r_1 = FwfwfWffWfwfwFww$
$r_2 = FFwFFwwFwfwfwFww$

$X_3 =$ 111000000001000110 011011101101011000 111101011110001100
111111100110110000 0000100010100010

X_3 quasi-relators:
$r_1 = FFwfwFFwwFWFwFWfWFWffWFWfWFwFWFww$
$r_2 = FFwfwFwfWfwfWWfwfWfwFwfwFFwwFWFww$

$X_{4a} =$ 111000000001000110 011001001111101010 011110110110111101
100011111110110110 10000111101

$X_{4b} =$ 111000000001000110 011001001111101010 111110010110011101
000011011110010110 00000101101

X_4 quasi-relators:
$r_1 = FFwfwFwfWfwfWfwFwfwFFwwFWFwFWFww$
$r_2 = FFwfwFwfwFFwwFWFwFWfWFWfWFwFWFww$

$X_{5a} =$ 001000110001110111 001111000101111111 101111100111001111
000001111011110111 1

$X_{5b} =$ 001001110000110110 001110000100111110 101110100110001110
000000111010110110 1

X_5 quasi-relators:
$r_1 = FwFWFwFwfwfWfwfw$
$r_2 = FwfwfWfWFWfWfwfw$

$X_{6a} =$ 111000000001000111 111111110101001111 011111010111111111
110001001011000111 0

$X_{6b} =$ 111001000000000110 111110110100001110 011110010110111110
110000001010000110 0

X_6 quasi-relators:

$r_1 = FWFwFWfWFwFWFwfw$

$r_2 = FWFwfwFwfWfwFwfw$

Proof: The proof follows along the lines presented in Remark 1.17. Two computer files contain the data needed for the proof. The first computer file describes the partition of $\mathcal{W}$ into sub-boxes and attaches an integer to each such sub-box, and the second file, called "conditionlist" is an ordered list of conditions and killerwords. The integer associated to a sub-box in the first file describes the numbered condition/killerword from conditionlist that will eliminate the sub-box in question (other than the X_i). A computer program named *verify* shows that the conditions and killerwords in question actually do kill off their associated sub-boxes (see Chapter 5 for more details). This computer program addresses the issues of Remark 1.20. The code for *verify* is available at the *Annals* web site.

In addition, a mild modification of *verify* showed that the listed words were quasi-relators for the given sub-boxes. ∎

Corollary 1.29: If δ is a shortest geodesic in N, a closed orientable hyperbolic 3-manifold, then

i) either tuberadius(δ) $> \ln(3)/2$ or $\exp(\text{length}(\delta)) \in \mathcal{L}(X_k)$ for some $k \in 0, \ldots, 6$ where $\mathcal{L}(X_k)$ denotes the range of L' values in the sub-box X_k. ∎

ii) either tuberadius(δ) $> \ln(3)/2$ or tuberadius(δ) $= \text{Re}(D)/2$ where $\exp(D) \in \mathcal{D}(X_k)$ for some $k \in 0, \ldots, 6$ and $\mathcal{D}(X_k)$ denotes the range of D' values in the sub-box X_k. ∎

Table 1.1: Exceptional Boxes in (L', D', R') Co-ordinates in $\mathcal{W}$; truncated values

X_0

$$l'_{min} = -0.84065 \quad l'_{max} = -0.84060 \quad t'_{min} = -2.13726 \quad t'_{max} = -2.13722$$

$$d'_{min} = -0.84064 \quad d'_{max} = -0.84059 \quad b'_{min} = -2.13729 \quad b'_{max} = -2.13722$$

$$r'_{min} = 0.999979 \quad r'_{max} = 1.000022 \quad a'_{min} = -0.00006103 \quad a'_{max} = 0.00006103$$

X_1

$$l'_{min} = -1.34852 \quad l'_{max} = -1.34831 \quad t'_{min} = -2.66102 \quad t'_{max} = -2.66072$$

$$d'_{min} = -0.54334 \quad d'_{max} = -0.54315 \quad b'_{min} = -2.85877 \quad b'_{max} = -2.85849$$

$$r'_{min} = 0.90390 \quad r'_{max} = 0.90408 \quad a'_{min} = -1.47167 \quad a'_{max} = -1.47143$$

X_2

$$l'_{min} = -1.78701 \quad l'_{max} = -1.78527 \quad t'_{min} = -2.27253 \quad t'_{max} = -2.27130$$

$$d'_{min} = -1.07428 \quad d'_{max} = -1.07273 \quad b'_{min} = -2.71846 \quad b'_{max} = -2.71736$$

$$r'_{min} = 0.74163 \quad r'_{max} = 0.74301 \quad a'_{min} = -1.52929 \quad a'_{max} = -1.52832$$

X_3

$$l'_{min} = 0.58117 \quad l'_{max} = 0.58160 \quad t'_{min} = -3.31221 \quad t'_{max} = -3.31190$$

$$d'_{min} = 1.15644 \quad d'_{max} = 1.15683 \quad b'_{min} = -2.75628 \quad b'_{max} = -2.75573$$

$$r'_{min} = 1.40420 \quad r'_{max} = 1.40454 \quad a'_{min} = -1.17968 \quad a'_{max} = -1.17919$$

X_4

$$l'_{min} = 0.33321 \quad l'_{max} = 0.33495 \quad t'_{min} = -3.31959 \quad t'_{max} = -3.31898$$

$$d'_{min} = 0.97739 \quad d'_{max} = 0.97817 \quad b'_{min} = -2.82533 \quad b'_{max} = -2.82478$$

$$r'_{min} = 1.35413 \quad r'_{max} = 1.35482 \quad a'_{min} = -1.22558 \quad a'_{max} = -1.22460$$

X_5

$$l'_{min} = -1.37984 \quad l'_{max} = -1.37810 \quad t'_{min} = -2.53706 \quad t'_{max} = -2.53460$$

$$d'_{min} = -1.37967 \quad d'_{max} = -1.37657 \quad b'_{min} = -2.53650 \quad b'_{max} = -2.53430$$

$$r'_{min} = 0.99989 \quad r'_{max} = 1.00265 \quad a'_{min} = -0.001953 \quad a'_{max} = 0.001953$$

X_6

$$l'_{min} = 1.37810 \quad l'_{max} = 1.37984 \quad t'_{min} = -2.53706 \quad t'_{max} = -2.53460$$

$$d'_{min} = 1.37657 \quad d'_{max} = 1.37967 \quad b'_{min} = -2.53650 \quad b'_{max} = -2.53430$$

$$r'_{min} = 0.99989 \quad r'_{max} = 1.00265 \quad a'_{min} = -0.001953 \quad a'_{max} = 0.001953$$

Table 1.2: Exceptional Boxes in (L, D, R) Co-ordinates in $\mathcal{P}$

$\mathcal{R}_0$

$$l_{min} = 0.8314 \quad l_{max} = 0.8315 \quad t_{min} = -1.9456 \quad t_{max} = -1.9455$$

$$d_{min} = 0.8314 \quad d_{max} = 0.8315 \quad b_{min} = -1.9456 \quad b_{max} = -1.9455$$

$$r_{min} = -0.00002051 \quad r_{max} = 0.00002267 \quad a_{min} = -0.00006105 \quad a_{max} = 0.00006105$$

$\mathcal{R}_1$

$$l_{min} = 1.0928 \quad l_{max} = 1.0931 \quad t_{min} = -2.0399 \quad t_{max} = -2.0397$$

$$d_{min} = 1.0680 \quad d_{max} = 1.0682 \quad b_{min} = -1.7587 \quad b_{max} = -1.7585$$

$$r_{min} = 0.5463 \quad r_{max} = 0.5465 \quad a_{min} = -1.0201 \quad a_{max} = -1.0198$$

$\mathcal{R}_2$

$$l_{min} = 1.0608 \quad l_{max} = 1.0617 \quad t_{min} = -2.2375 \quad t_{max} = -2.2366$$

$$d_{min} = 1.0720 \quad d_{max} = 1.0727 \quad b_{min} = -1.9473 \quad b_{max} = -1.9466$$

$$r_{min} = 0.5298 \quad r_{max} = 0.5308 \quad a_{min} = -1.1193 \quad a_{max} = -1.1182$$

$\mathcal{R}_3$

$$l_{min} = 1.2126 \quad l_{max} = 1.2129 \quad t_{min} = -1.3972 \quad t_{max} = -1.3969$$

$$d_{min} = 1.0947 \quad d_{max} = 1.0951 \quad b_{min} = -1.1736 \quad b_{max} = -1.1733$$

$$r_{min} = 0.6063 \quad r_{max} = 0.6067 \quad a_{min} = -0.6988 \quad a_{max} = -0.6984$$

$\mathcal{R}_4$

$$l_{min} = 1.2046 \quad l_{max} = 1.2050 \quad t_{min} = -1.4708 \quad t_{max} = -1.4702$$

$$d_{min} = 1.0949 \quad d_{max} = 1.0953 \quad b_{min} = -1.2378 \quad b_{max} = -1.2374$$

$$r_{min} = 0.6019 \quad r_{max} = 0.6027 \quad a_{min} = -0.7357 \quad a_{max} = -0.7349$$

$\mathcal{R}_5$

$$l_{min} = 1.0595 \quad l_{max} = 1.0606 \quad t_{min} = -2.0694 \quad t_{max} = -2.0683$$

$$d_{min} = 1.0591 \quad d_{max} = 1.0604 \quad b_{min} = -2.0694 \quad b_{max} = -2.0680$$

$$r_{min} = -0.0001069 \quad r_{max} = 0.002654 \quad a_{min} = -0.001954 \quad a_{max} = 0.001954$$

$\mathcal{R}_6$

$$l_{min} = 1.0595 \quad l_{max} = 1.0606 \quad t_{min} = -1.0733 \quad t_{max} = -1.0722$$

$$d_{min} = 1.0591 \quad d_{max} = 1.0604 \quad b_{min} = -1.0736 \quad b_{max} = -1.0722$$

$$r_{min} = -0.0001069 \quad r_{max} = 0.002654 \quad a_{min} = -0.001954 \quad a_{max} = 0.001954$$

Remarks 1.30: i) The values in Table 1.1 are only approximations of actual values which can be computed as in Remark 1.26iii.

ii) The values in Table 1.2 correspond to boxes in $\mathcal{P}$ which contain the natural log of boxes in Table 1.1 (here we use the true co-ordinates of the boxes, not just the approximation-by-truncation co-ordinates). For example, the rectangle in $\mathbf{C}$ determined by $l_{min}, l_{max}, t_{min}, t_{max}$ for $\mathcal{R}_0$ contains ln(the rectangle in $\mathbf{C}$ determined by $l'_{min}, l'_{max}, t'_{min}, t'_{max}$ for X_0). $\blacksquare$

iii) In Chapter 3 we show that $X_0 \cap \mathcal{T}$ is a single parameter. This parameter has $L = D = 0.8314429..... - (1.9455307...)i$ and $R = 0$.

iv) Recall that $\ln(3) = 1.10986...$. Thus we see that all boxes but $\mathcal{R}_0$ lie close to the boundary of the parameter space $\mathcal{P}$. Boxes $\mathcal{R}_3, \mathcal{R}_4$ are particularly close to the edge.

Conjecture 1.31: Each sub-box $X_i, 0 \le i \le 6$ contains a unique element s_i of S. Further, if $\{G_i, f_i, w_i\}$ is the marked group associated to s_i then $N_i = \mathbf{H}^3/G_i$ is a closed hyperbolic 3-manifold with the following properties

i) N_i has fundamental group $< f, w; r_1(X_i), r_2(X_i) >$, where $r_1(X_i),\ r_2(X_i)$ are the quasi-relators associated to the box X_i.

ii) N_i has a heegaard genus 2 splitting realizing the above group presentation.

iii) N_i nontrivially covers no manifold.

iv) N_6 is isometric to N_5.

v) If (L_i, D_i, R_i) is the parameter in $\mathcal{T}$ corresponding to s_i, then L_i, D_i, R_i are related as follows.

For $X_0,\ X_5,\ X_6:\quad L = D,\quad R = 0.$

For $X_1,\ X_2,\ X_3,\ X_4:\quad R = L/2.$

Remarks 1.32: i) Experimentally, at some point in the sub-box X_i under consideration the quasi-relators are actually relators.

ii) In Chapter 3, we prove Conjecture 1.31 for X_0. Recently, K. Jones and A. Reid have made further progress on Conjecture 1.31 (see [JR]). Among other things, they construct arithmetic hyperbolic 3-manifolds associated to each sub- box but X_3 and prove that the manifolds so constructed for X_5 and X_6 are isometric. They are able to prove uniqueness in the case of X_0 and the techniques they use generalize to the other arithmetic hyperbolic 3- manifolds, but the analysis is more complicated and the question of uniqueness remains open in these cases.

iii) According to *Heegaard* [Be], for each i, there exists a closed 3-manifold B_i of Heegaard genus 2 such that $\pi_1(B_i) =< f, w; r_1(X_i), r_2(X_i) >$ and this group presentation arises naturally from the Heegaard splitting.

iv) Experimental evidence suggests that the manifolds associated to X_5 and X_6 are both isometric to the Weeks census manifold $s479(-3, 1)$. Also experimentally, the manifold associated with X_2 is $s778(-3, 1)$ and the manifold associated with X_1 is $v2678(2, 1)$.

v) Except for Vol3, R. Riley's program *Poincaré* (see [Ri]) was the first to show (experimentally) that there is a closed orientable (hyperbolic) 3-manifold associated to each box. It provided group presentations, which presumably could be shown to be those of the above.

vi) Table 1.2, Conjecture 1.31, and Remark 1.32ii indicate that Proposition 1.11ii is somewhat close to being sharp.

The following conjecture is a succinct, though slightly weaker form of Conjecture 1.31.

Conjecture 1.33: If δ is a shortest geodesic in a closed orientable hyperbolic 3-manifold N, then either tuberadius$(\delta) > \ln(3)/2$ or N is one of six exceptional manifolds.

Remark 1.34: Here we outline our method for finding a decomposition of the initial box $\mathcal{W}$ into sub-boxes each with a condition/killerword that kills off the entire associated sub-box.

A simple algorithm for finding a killerword for a region is as follows. Work with a set of words to consider, initialized to the null word. At each step, remove the oldest word from the set, and test to see if that word is a killerword. If it isn't, put the word back into the set, concatenated with each of the generators and their inverses. Eventually, this algorithm will enumerate all words, and so, if there is a killerword, the algorithm will eventually find it. In practice, there are two problems with this approach: there is no provision for the possibility that no killerword exists for the region under consideration, and the time to find a word of length n grows exponentially.

When we also take into account the possibility of subdividing the box, getting an answer in finite time will be possible; but the search is in practice very expensive. The most obvious way of speeding it up is to avoid the search entirely when feasible: a killerword works on a neighborhood of a region, and by testing killerwords found for nearby boxes, most of the time the search is not necessary.

Still, there are words of length as long as 44 that were considered, and testing all of the roughly 3^{44} combinations would be prohibitive on today's computers. In practice (due to a bug), the search algorithm used for most of the parameter space was no better than the brute-force method just described, but to find killerwords for the remaining regions, an improvement was needed. Rather than blindly selecting words in first-in-first-out order, the algorithm can rank the words under consideration based on a heuristic estimate of the likelihood of their being useful (a word is *useful* if it is a prefix of a killerword). We note first that short words tend to be better than long words, as they have fewer steps and less error. Second, we note that words with a large translation distance are given a bad ranking, for two reasons: they will need more generators appended before they get back to the small translation distance which is needed for a contradiction, and computations with those words introduce more error per step than computations with closer words.

This approach was an improvement, but wasn't finding enough killerwords in the regions around X_3 and X_4. Further investigation showed that the algorithm was getting stuck on an identity: once it found an identity, it would consider only words which started with that identity, and ignored all of the other words. To fix this problem, a "diversity" heuristic was introduced, to give special consideration to unlikely but unusual words.

To prevent the search from running forever, the search is temporarily abandoned after some number of steps, and re-done with twice as many steps every time the number of descendant boxes doubles. This way, the search could run forever, but only if the subdivision process runs forever. This merged process of alternately searching and subdividing we call the *decomposition algorithm.*

The decomposition algorithm went through several revisions; at each stage of the revision process, the algorithm effectively increased the extent to which killerwords found for one region were used to kill other regions. The first attempt—used to determine the feasibility of the whole effort— iterated over regions in depth-first order, performing the search as described above. At that stage, it became evident that the search process, as opposed to the sub-division process, was consuming nearly all of the compute time, and so the second version iterated over regions in breadth-first order, and once it found a killerword, tried to use that word on all adjacent regions.

The breadth-first version was used to analyze the entire parameter space, although it skipped some parts due to various bugs; the search heuristic was replaced once, and there was considerable human input to tell the search about particularly difficult killerwords, or to tweak its search parameters (length, and weightings in the heuristic).

The third stage of the revision process reduced the number of boxes by attempting all found killerwords in a large region (about a thousand boxes) on all boxes in the region. It did not do any searching, since it was provided with a list of killerwords known to work.

The final version was created when the bugs in evaluation were brought to light, and the existing killerword tree was found to be insufficient. It used the list of killerwords used for the entire tree, and some statistics about the number of subdivisions required in order for a given word to kill a particular sub-box, and evaluated each word on each box. Whenever a word was evaluated, a kind of triage was used to determine whether that word was likely to kill the box in question, likely to kill any of its n^{th} generation descendants, or unlikely to kill any descendants of the box; the answer to that heuristic either allowed more detailed evaluation (with the error term included), deferred further evaluation until the box had been subdivided n more times, or excluded that word from further consideration on any descendant of the box. With these heuristics, this program wound up evaluating on average about 10 of the roughly 13200 words per box, and was able to construct the tree consisting of the decomposition into sub-boxes with associated conditions/killerwords.

We mention that the bugs, complexity, and frequent changes in the search programs are irrelevant to the accuracy of the verification. In fact, we shielded the verification programs from internal issues related to the searching programs. Given a putative decomposition of the parameter space into sub-boxes with associated conditions/killerwords the program *verify* simply checks whether this decomposition with conditions/ killerwords works.

Preview 1.35: In Remark 1.17, we mentioned that we use first-order Taylor approximations, with Remainder term, to show that a killerword which eliminates a point $x \in \mathcal{W}_i$ eliminates all of $\mathcal{W}_i$. The computational object we constructed to carry out these Taylor approximations is called an *AffApprox*. In the parameter space $\mathcal{W}$, all functions analyzed via Taylor approximations in this way are built up from the operations $+,\ -,\ \times,\ /,\ \sqrt{\ }$. We prove combination formulas for these operations, which show how the Taylor approximations (including the Remainder term) change when one of these operations is applied to two AffApprox's. This is carried out in Chapter 6.

To ensure that all of our computer calculations are rigorous, we use a round-off error analysis. Typically, this is done by using interval arithmetic on floating-point numbers. Instead, we introduce round-off error at the level of AffApprox's and incorporate the round-off error into the Remainder term. The main reason for this additional complexity is to get more accuracy in our calculation of AffApprox's, which allows us to analyze substantially fewer boxes. Further, the individual computations are faster. This is all carried out in Chapters 7 and 8.

Chapter 2: The Corona Insulator Family

The upshot of Proposition 1.28 is that if a closed orientable hyperbolic 3-manifold has a shortest geodesic which does not have an embedded $\ln(3)/2$ tube then the parameters for its associated marked group(s) (G, f, w) must be in one of the boxes $X_0, X_1, \ldots, X_6$ listed in that proposition. Nonetheless—as we shall see in this Chapter and the next— such manifolds have non-coalescable insulator families about their shortest geodesics, although they might not be Dirichlet insulator families. A review of the theory of insulators in hyperbolic 3-manifolds is provided in Appendix I.

In this Chapter, we describe a new insulator family $\{\kappa_{ij}\}$ called the Corona insulator, and we describe a condition sufficient for this family to be non-coalescable—a condition which is weaker than the tuberadius$(\delta) > \ln(3)/2$ sufficient condition for the Dirichlet insulator family.

The reason Dirichlet insulator families for geodesics with solid tubes of radius greater than $\ln(3)/2$ are non-coalescable is that the amount of visual angle taken up by the various insulators is less than 120 degrees, and thus there is no chance for tri-linking to occur. The visual angle (measured at one axis) for a member of the Dirichlet insulator family associated to two axes depends only on the real distance between the two axes.

In contrast, the visual angle for a member of the Corona insulator family associated to two axes depends on the complex distance between the two axes. We now define this function $\mathcal{C}$ of complex distance, and name it the *Corona* function. After that we give a precise definition of the visual angle function, and prove that the Corona function is the proper visual angle function for the Corona insulator family.

Definition 2.1: Let $\mathcal{C} : (0,\infty) \times [-\pi,\pi] \to (0,\pi)$ be defined by

$$\mathcal{C}(u,v) = \text{Abs}(\text{Im}(\text{Arccosh}(1 - \frac{4}{1 \pm \cosh(u+iv)})))$$

where $\pm$ is positive for $-\pi/2 \leq v \leq \pi/2$ and negative otherwise. The branch of Arccosh is chosen so that the values of the Corona function lie between 0 and π.

Figure 2.1: The 102 to 120 degree contours for the Corona function $\mathcal{C}(u,v)$ (for example, the 120 degree contour corresponds to where the Corona function takes on value $2\pi/3$).

In the following definition, it is helpful to imagine the geodesic σ as being the z-axis in the upper-half-space model of $\mathbf{H}^3$.

Definition 2.2: If $\sigma \subset \mathbf{H}^3$ is a geodesic, then $S^2_\infty - \partial\sigma$ can be parametrized by $S^1 \times \mathbf{R}$, where $\mathbf{R}$ represents the set of real numbers. Each $x \times \mathbf{R}$ lies in the ideal boundary of a hyperbolic halfplane bounded by σ. Two such lines $x \times \mathbf{R}$, $y \times \mathbf{R}$ are at distance θ in the S^1 factor if they meet at $\partial\sigma$ at angle θ. Consider a region R in $S^2_\infty - \partial\sigma$, and define $\text{visualangle}_\sigma(R) = \theta \in [0, 2\pi]$ as follows. Let θ be the infimum of lengths of closed subintervals J of S^1 such that $R \subset J \times \mathbf{R}$, if there is such a subinterval, and let $\theta = 2\pi$ if there is no such subinterval.

Proposition 2.3: Let δ_i, δ_j be disjoint oriented geodesics in $\mathbf{H}^3$. Then there exists a smooth simple closed curve κ_{ij} in S^2_∞ separating $\partial\delta_i$ from $\partial\delta_j$ such that for $k \in \{i,j\}$, $\text{visualangle}_{\delta_k}(\kappa_{ij}) = \mathcal{C}(\text{distance}(\delta_i,\delta_j))$.

Proof: Let P be the orthocurve from δ_i to δ_j. Consider the half-plane with boundary δ_i containing P, and the half-plane with boundary δ_j containing P. Allow these half-planes to expand into wedges at the same rate. (A *wedge* is a closed set in $B^3 = \mathbf{H}^3 \cup S^2_\infty$ bounded by two hyperbolic halfplanes which meet along a common geodesic.) At first, the four half-planes that bound these wedges intersect in $\mathbf{H}^3$, but at some angle θ these half-planes intersect only at infinity (that is, in S^2_∞). By reasons of symmetry they intersect in two or four points (four points of intersection occur when $\mathrm{Im}(\mathrm{distance}(\delta_i, \delta_j))$ is $\pi/2$ or $-\pi/2$).

Let A_i, A_j be the wedges which exist at angle θ. Let $T_k = A_k \cap S^2_\infty$ for $k \in \{i, j\}$. Let κ_{ij} be a simple closed curve in $T_i \cap T_j$ which separates $\partial\delta_i$ from $\partial\delta_j$. By construction, for $k \in \{i, j\}$, $\mathrm{visualangle}_{\delta_k}(\kappa_{ij}) = \theta$, and θ is in $[0, \pi]$. See Figure 2.2 (and compare with Figure 2.3).

Figure 2.2: The lightly shaded region is $T_i \cap T_j$, while the darkly shaded region is $S^2_\infty - T_j$.

To complete the proof of the Proposition, we now show that $\theta = \mathcal{C}(\mathrm{distance}(\delta_i, \delta_j))$. To do this, we use hyperbolic trigonometry on a degenerate right-angled hexagon in $\mathbf{H}^3$. Following [F], a degenerate right-angled hexagon is a 5-tuple of oriented geodesics $S_1, \cdots, S_5$ in $\mathbf{H}^3$ such that S_i is orthogonal to S_{i+1} and S_1 and S_5 limit at a common point S_0 at infinity. These oriented geodesics give rise to complex numbers $\sigma_0, \sigma_2, \sigma_3, \sigma_4$ representing signed edge lengths. For $k \in \{2, 3, 4\}$ $\sigma_k = d_{S_k}(S_{k-1}, S_{k+1})$ where a $d_{S_k}(S_{k-1}, S_{k+1})$ translation of $\mathbf{H}^3$ along the oriented geodesic S_k takes the oriented geodesic S_{k-1} to the oriented geodesic S_{k+1} (see definition 1.4). The remaining (degenerate) edge length is given by $\sigma_0 = 0$ if the axes S_1 and S_5 either both point into S_0 or both point out of S_0; otherwise $\sigma_0 = \pi i$. By [F; pg. 83] we have the following Hyperbolic Law of Cosines.

$$\cosh(\sigma_0) = \cosh(\sigma_2)\cosh(\sigma_4) + \sinh(\sigma_2)\sinh(\sigma_4)\cosh(\sigma_3).$$

We work in the upper-half-space model of hyperbolic 3-space, and normalize so that the ortholine from δ_j to δ_i is $B_{(0;\infty)}$ (thus δ_i intersects $B_{(0;\infty)}$ above δ_j), while the oriented axis δ_i is $B_{(-1;1)}$. $B_{(0;\infty)}$ will be S_3 while the oriented geodesics δ_i and δ_j will be S_2 and S_4, respectively.

Let, $u + iv = \mathrm{distance}(\delta_i, \delta_j)$. If $-\pi/2 < v < 0$ then the intersection points at infinity occur in the second quadrant and the fourth quadrant. See Figure 2.3a. For convenience, we work with the point in the second quadrant and send (unique) perpendiculars from it to the geodesics δ_i and δ_j. The perpendicular to δ_i will be oriented towards δ_i and then denoted S_1, while the perpendicular to δ_j will be oriented away from δ_j and then denoted S_5. The intersection point at infinity (in the second quadrant) is S_0.

Figure 2.3: Two versions of the degenerate right-angled hexagon.

This is the proper set-up for applying the (degenerate) Hyperbolic Law of Cosines (see Figure 2.3b). Note that $\sigma_3 = -(u + iv)$, and $\sigma_0 = i\pi$. By symmetry $\sigma_2 = \sigma_4 = (\alpha + i\beta)/2$ where $(\alpha + i\beta)/2$ is $\mathrm{distance}(S_1, S_3)$. Plugging into the Law of Cosines, using a half-angle formula ($\cosh(2z) = 2\cosh^2(z) - 1 = 2\sinh^2(z) + 1$), solving for $\cosh(\alpha + i\beta)$, and taking the Arccosh, we get the desired result. Note that the visual angle in this set-up is $-\beta$, thus necessitating taking the absolute value.

When $0 < v < \pi/2$ our 2 intersection points occur in the first and third quadrants, and we carry out the same procedure. This time S_2 and S_4 are traversed in the direction opposite to their orientations (the attendant changes in sign drop out though). In this case, the visual angle is β.

The special cases $v = -\pi/2,\ 0,\ \pi/2$ follow similarly.

The cases $-\pi \leq v \leq -\pi/2$ and $\pi/2 \leq v \leq \pi$ reduce to the previous cases after adding or subtracting π. The formula in Definition 2.1 is then obtained after noting that $\cosh(z \pm i\pi) = -\cosh(z)$. ∎

Definition 2.4: Let δ be a simple closed geodesic in the closed orientable hyperbolic 3-manifold N. Let $\{\delta_i\}_{i\geq 0}$ be the lifts of δ to $\mathbf{H}^3$. For each $\pi_1(N)$-orbit of unordered pairs (δ_i, δ_j) choose a representative where $i = 0$. If $\text{Redistance}(\delta_0, \delta_i) \leq \ln(3)/2$, then let κ_{0j} be a smooth simple closed curve in S^2 separating $\partial\delta_0$ from $\partial\delta_j$ such that for $k \in \{0, j\}$, $\text{visualangle}_{\delta_k}(\kappa_{0j}) = \mathcal{C}(\text{distance}(\delta_0, \delta_j))$. If $\text{Redistance}(\delta_0, \delta_i) > \ln(3)/2$, then let κ_{0j} be the Dirichlet insulator, i.e. the boundary of the geodesic plane orthogonally bisecting the orthocurve between δ_0, δ_j.

In either case, extend the collection $\pi_1(N)$-equivariantly to a family $\{\kappa_{ij}\}$ defined for all i, j. This is the *Corona family* for δ.

Lemma 2.5: The Corona family $\{\kappa_{ij}\}$ is

i) an insulator family for δ

ii) noncoalescable if $\max\{\mathcal{C}(\delta_0, \delta_j) \mid j > 0\} < 2\pi/3$.

Proof: We check that $\{\kappa_{ij}\}$ satisfies the various conditions of Definitions A.1 and A.2.

i) By construction, κ_{ij} separates $\partial\delta_i$ from $\partial\delta_j$ and $\{\kappa_{ij}\}$ is $\pi_1(N)$-equivariant. For $k \in \{i, j\}$, δ_k-visualangle$(\kappa_{ij}) < \pi$, and so $\{\kappa_{ij}\}$ satisfies the convexity condition.

Modulo the natural action of $\pi_1(N)$ on κ_{ij}, there are only finitely many insulators κ_{ij} which are not Dirichlet insulators. Therefore, for fixed i, there exist only finitely many κ_{ij} such that $\text{diam}(\kappa_{ij}) > \epsilon$. This establishes local finiteness. ∎

ii) No trilinking follows immediately from the "less than $2\pi/3$" condition. ∎

Definition 2.6: If δ is a simple closed geodesic in the hyperbolic 3-manifold N, define $\text{maxcorona}(\delta) = \max\{\mathcal{C}(\delta_0, \delta_j) \mid j > 0\}$

Remarks 2.7: i) It seems possible that the Dirichlet insulator family associated to a geodesic $\delta \in N$ may be non-coalescable, while a Corona insulator family is coalescable and conversely (Corona family non-coalescable while Dirichlet family coalescable).

ii) If $\text{tuberadius}(\delta) > \ln(3)/2$, then the Dirichlet insulator and Corona insulator families coincide and by Lemma 5.9 of [G] they are non-coalescable.

iii) The corona family is not uniquely defined for there is some choice in constructing the κ_{0j}'s.

Proposition 2.8: Let δ be a shortest geodesic in the closed orientable hyperbolic 3-manifold N. Then either a Corona insulator family is noncoalescable or there exists a marked group $\{G, f, w\}$ where G is a subgroup of $\pi_1(N)$ and the parameter associated to $\{G, f, w\}$ lies in the box $X_0 = X_{0a} \cup X_{0b} \subset \mathcal{W}$.

Proof: Recall that $\mathcal{L}(X_k)$ denotes the range of L' values in the sub-box X_k, and $\mathcal{D}(X_k)$ denotes the range of D' values in the sub-box X_k.

Let δ be a shortest geodesic in N, a closed orientable hyperbolic 3-manifold. By Corollary 1.29, Lemma 2.5, and Remark 2.7ii, either δ has a non-coalescable insulator family or length(δ) lies in $\mathcal{L}(X_k)$ for some $k \in \{0, 1, \ldots, 6\}$ and maxcorona$(\delta) \geq 2\pi/3$. Let $\{G, f, w\}$ be a marked group where G is a subgroup of $\pi_1(N)$, f corresponds to δ and has axis $B_{(0;\infty)}$, and g maximizes $\mathcal{C}(g(B_{(0;\infty)}), B_{(0;\infty)})$. We will show that if $\mathcal{C}(g(B_{(0;\infty)}), B_{(0;\infty)}) \geq 2\pi/3$, then the parameter associated to $\{G, f, g\}$ lies in the sub-box X_0.

In fact we will show that if $\{G, f, g\}$ is any marked group where G is a torsion-free parabolic-free subgroup of $\text{Isom}_+(\mathbf{H}^3)$, f is a length-minimizing loxodromic element with axis $B_{(0;\infty)}$, and g maximizes $\mathcal{C}(g(B_{(0;\infty)}), B_{(0;\infty)})$ where $\mathcal{C}(g(B_{(0;\infty)}), B_{(0;\infty)}) \geq 2\pi/3$, then the parameter $\beta = (L_\beta, D_\beta, R_\beta) \in \exp^{-1}(X_0)$.

It follows as in the second paragraph of the present proof that $L_\beta \in \mathcal{L}(X_k)$ for some $k \in \{0, 1, 2, 3, 4, 5, 6\}$. Also D_β is subject to nontrivial constraint. That is, if $L_\beta \in \mathcal{L}(X_j)$ for $j \in \{1, 2, 3, 4, 5, 6\}$, then D_β must lie in the decorated region of Figure 2.4. This follows for two reasons. First, Redistance$(g(B_{(0;\infty)}), B_{(0;\infty)}) \geq \min_d(X_j) > 1.059$ where $\min_d(X_j)$ is the minimal d value associated to X_j, and 1.059 is a computed lower bound for $\min_d(X_j)$ which works for all $j \in \{1, 2, 3, 4, 5, 6\}$ as in Table 1.2. (The fact that the $\mathcal{L}(X_k)$'s are disjoint implies that if $\beta \in S \cap X_j$, then tuberadius$(\delta) \geq \min_d(X_j)/2$, where δ corresponds to the element f.) Second, $\mathcal{C}(g(B_{(0;\infty)}), B_{(0;\infty)}) \geq 2\pi/3$ implies that D_β must lie in the decorated region of Figure 2.4.

Figure 2.4: The 120-degree contour for the Corona function $\mathcal{C}(u, v)$, and a decorated region which corresponds to Corona greater than or equal to 120 degrees and u value greater than or equal to 1.059.

Our proof is now similar to the proof of Proposition 1.28. We partition the initial box $\mathcal{W}$ into sub-boxes $\mathcal{W}_i$ and eliminate $\mathcal{W}_i$ if any of the following conditions hold (for clarity, the conditions have been translated into "pre-exponentiated" form).

a) There exists no $\beta \in \mathcal{W}_i$ such that length$(f_\beta) \in \mathcal{L}(X_k)$.

b) $\mathcal{W}_i$ has some L values in (exactly one) $\mathcal{L}(X_k)$ but there exists no $\beta \in \mathcal{W}_i$ such that distance $(w_\beta(B_{(0;\infty)}), B_{(0;\infty)}) \in$ decorated region.

c) There exists a killerword h in f, w, f^{-1}, w^{-1} such that Relength$(h_\beta) <$ Relength(f_β) and $h_\beta \neq$ id for all $\beta \in \mathcal{W}_i$.

d) There exists a killerword h in f, w, f^{-1}, w^{-1} such that

$$\mathcal{C}(\text{distance}(h_\beta(B_{(0;\infty)}), B_{(0;\infty)})) > \mathcal{C}(\text{distance}(w_\beta(B_{(0;\infty)}), B_{(0;\infty)}))$$

and $h_\beta(B_{(0;\infty)}) \neq B_{(0;\infty)}$ for all $\beta \in \mathcal{W}_i$.

We have two files that contain the decomposition of $\mathcal{W}$ into sub-boxes and associated conditions/killerwords. The program *corona* checks that these files do indeed eliminate all of $\mathcal{W} - X_0$. *Corona* analyzes the cases $k = 1, \ldots, 6$ all at once. ∎

Remarks 2.9: i) The proof of Proposition 2.8 requires working in a considerably smaller parameter space than that of Proposition 1.28. Condition a) implies that the parameter space is "$(2+\epsilon)$-complex dimensional" and condition b) implies that one of the parameters is greatly constrained. This suggests why it took so much longer to come up with the partition and the associated killerwords for Proposition 1.28. In fact, it took roughly 1500 CPU days to find the partition and the associated killerwords for Proposition 1.28, versus roughly 2 CPU days for Proposition 2.8. Here, the term "CPU day" refers to 24 hours of running an SGI Indigo 2 workstation with an R4400 chip, and the estimate of 1500 CPU days refers to 15 to 20 such machines running 80 to 90 percent of the time for 3 to 4 months.

ii) We took pains to make *corona* as similar to *verify* as we could, thereby lessening the amount of analysis needed to show the veracity of *corona*.

iii) When working with exponentiated co-ordinates (that is, in $\mathcal{W}$ rather than $\mathcal{P}$) the Corona function changes as follows. Let $X = \exp(\alpha + i\beta)$ and $U = \exp(u + iv)$, then the Corona function when $-\pi/2 \leq v \leq \pi/2$ (so that the positive choice in $\pm$ is made)

$$\cosh(\alpha + i\beta) = 1 - \frac{4}{1 + \cosh(u + iv)}$$

becomes

$$\frac{X + X^{-1}}{2} = 1 - \frac{4}{1 + (U + U^{-1})/2}$$

It is a pleasant exercise to solve this, and we find that

$$X = \frac{(U^2 - 6U + 1) \pm 4(U - 1)\sqrt{-U}}{(U + 1)^2}.$$

The two answers are reciprocals, which implies their associated arguments are opposites. We choose $+$ or $-$ so that the argument is positive.

For $-\pi \leq v \leq -\pi/2$ and $\pi/2 \leq v \leq \pi$ (so that the negative choice of $\pm$ is made) we find (replace U by $-U$ in the above formula) that

$$X = \frac{(U^2 + 6U + 1) \pm 4(U + 1)\sqrt{U}}{(U - 1)^2}.$$

In *corona*, the exponentiated version of the Corona function is the function*horizon(ortho)*, which takes in $U =$ *ortho* and computes the associated X value. β, the argument of X, is implicitly obtained in the function *larger-angle*.

iv) It is possible that by working purely in the context of the Corona function, rather than first working with Redistance and attempting to prove Proposition 1.28, the computer proof can be simplified. We started this project with the naive idea that perhaps Vol3 was the only manifold whose shortest geodesic did not have a ln(3)/2 tube. The remarkable fact that this naive idea is almost correct accounts for the fact that a proof of Theorem 0.2 can be obtained with only the mild extra effort detailed in this chapter and the next.

Chapter 3: Vol3

In Chapter 1, we showed how a certain 3-complex-dimensional parameter space can be used to attack the question of whether all closed, oriented hyperbolic 3-manifolds have a tube of radius ln(3)/2 around a shortest geodesic and hence have non-coalescable insulator families. In particular, Proposition 1.28 showed that a hyperbolic 3-manifold has such a tube unless its fundamental group contains a marked group associated to a point in the seven exceptional regions $X_0, ..., X_6$.

In Chapter 2, we showed that any shortest geodesic in a closed oriented hyperbolic 3-manifold has a noncoalescable insulator family unless its fundamental group contains a marked group associated to a point in the region X_0. In this chapter, we show that the region X_0, or equivalently $\mathcal{R} = \exp^{-1}(X_0)$, is associated to a unique closed manifold and that manifold has a noncoalescable insulator family.

Proposition 3.1: The point $(\omega, \omega, 0)$ is the unique point in $\mathcal{T} \cap \mathcal{R}$; here $\omega = \ln((-1 - i * \sqrt{3})/2 - (-6 + 2 * i * \sqrt{3})^{(1/2)}/2) \approx 0.83144 - 1.94553 * i$. This point corresponds to Vol3, the closed hyperbolic 3-manifold of conjecturally third smallest volume. Further, if N is a closed orientable hyperbolic 3-manifold, either $N =$ Vol3 or maxcorona$(\delta) < 2\pi/3$ for δ a shortest geodesic in N.

Proposition 3.2: Any shortest geodesic in Vol3 satisfies the insulator condition.

Remark 3.3: Topologically, Vol3 is (3,1) surgery on manifold m007 in the census of cusped hyperbolic 3-manifolds (see [CHW]). It is also (-3,2) (-6,1) surgery on the left-handed Whitehead link, link 5^2_2 in the standard knot tables. The program *SnapPea* (see [W1]) gives an experimental proof that Vol3 is hyperbolic and that its volume is that of the regular ideal 3-simplex. A rigorous proof can be found in [JR] or [CGHN]. Previously, Hodgson-Weeks [HW1] had found an exact Dirichlet domain for Vol3, that is, the face pairings were expressible as explicit matrices with coefficients in a finite extension F of $\mathbf{Q}$ and they obtained equations in F for the various faces. See Remark 3.14.

Remark 3.4: Idea of Proof of Proposition 3.1: By the proof of Proposition 2.8, if N has maxcorona$(\delta) \geq 2\pi/3$, then it must have an (L, D, R) parameter in $\mathcal{R} \cap \mathcal{T}$. A geometric argument (Lemma 3.7) which utilizes Lemmas 3.5 and 3.6 shows that $R = 0$, and an algebraic argument (Lemmas 3.8, 3.9, 3.11, 3.13) shows that $L = D = \omega$, where $\exp(\omega)$ is a root of the polynomial $z^4 + 2z^3 + 6z^2 + 2z + 1$. This implies ([HW1] or [JR]) that Vol3 $= \mathbf{H^3}/G$ where G is the subgroup of $\pi_1(N)$ generated by f, w associated to the parameter $(L, D, R) = (\omega, \omega, 0)$. Therefore, N is covered by Vol3. Because Vol3 covers no 3-manifolds non-trivially ([JR]), it follows that $N =$ Vol3.

Lemma 3.5: Let $\mathcal{R} = \exp^{-1}(X_0)$. If $\alpha = (L, D, R) = (l + it, d + ib, r + ia) \in \mathcal{R} \cap \mathcal{T}$, then f_α, w_α satisfy the relations

i) $wFwfwfWfwf$

ii) $wFwfwwfwFw$

Here W denotes w^{-1}, and F denotes f^{-1}.

Proof: In i), ii) above and what follows below we supress the subscripts α. Because i), ii) are cyclic permutations of the quasi-relators r_2, r_1 corresponding to the X_0 box of Proposition 1.28, it follows that if $h = wFwfwfWfwf$ or $h = wFwfwwfwFw$, then Relength$(h) <$ Relength(f) throughout $\mathcal{R}$. Since $\alpha \in \mathcal{T}$, f is a shortest element and so $h = \text{id}$. ∎

Lemma 3.6: The following substitutions in Lemma 3.5 give rise to three sets of new relators.

a) In i), ii) exchange f and w (hence, exchange F and W).
b) In i), ii) exchange f and F.
c) In i), ii) exchange w and W.

Proof: a) First, a cyclic permutation of relator i) gives relator i) with f replaced by w and w replaced by f. Second, one readily obtains the relator $fWfwffwfWf$ from i), ii), because $fWfwf = (wFwfw)^{-1} = wfwFw = (fwfWf)^{-1}$ where the first and third equalities follow from i) and the second from ii).

b) Again it is routine to obtain b) from i), ii).

c) Conclusion c) follows from a) and b). ∎

Lemma 3.7: If $(L, D, R) \in \mathcal{R} \cap \mathcal{T}$, then $R = 0$.

Proof: Let $\{G, f, w\}$ be the marked group corresponding to the parameter (L, D, R). Figure 3.1 shows a schematic picture of geodesics $B_{(0;\infty)}$, $W(B_{(0;\infty)})$, $w(B_{(0;\infty)})$, $f(w(B_{(0;\infty)}))$, $f(W(B_{(0;\infty)}))$, where, in the figure we have, for convenience, abbreviated $B_{(0;\infty)}$ to B. Also, it shows the images of the orthocurve O from $W(B_{(0;\infty)})$ to $B_{(0;\infty)}$ after translation by w, f, and fw. Finally, it shows the orthocurves O_1 from $fW(B_{(0;\infty)})$ to $W(B_{(0;\infty)})$ and O_2 from $w(B_{(0;\infty)})$ to $fw(B_{(0;\infty)})$. Note that Figure 3.1 displays the situation where $\text{Re}(R) > 0$. It is also *a priori* possible that O_2 might intersect $w(B_{(0;\infty)})$ on the other side of $w(O)$. There are other similar possible inaccuracies.

Figure 3.1: Various geodesics around $B = B_{(0;\infty)}$ in $\mathbf{H^3}$.

$\sigma_1 = fwffw \in G$ sends geodesic $W(B_{(0;\infty)})$ to $fw(B_{(0;\infty)})$ and $\sigma_2 = wFFwF \in G$ sends the geodesic $fW(B_{(0;\infty)})$ to $w(B_{(0;\infty)})$. Now $\sigma_2^{-1}\sigma_1 = fWffWfwffw$ is a relator of G, since it is a cyclic permutation of relator ii) with f replaced by w and w replaced by f. Thus $\sigma_1 = \sigma_2$ and hence $\sigma_1(O_1) = O_2$.

Figure 3.1 gives rise to some right-angled hexagons which we will study. For a careful development of the theory of right-angled hexagons, see [F]; we give an abbreviated treatment here. A right-angled hexagon consists of a cyclically ordered 6-tuple of oriented geodesics $\lambda_1, \cdots, \lambda_6$ in $\mathbf{H^3}$ such that λ_i intersects $\lambda_{i+1} \pmod 6$ orthogonally. Each "edge" of the hexagon is labeled by the complex number $e_i = d_{\lambda_i}(\lambda_{i-1}, \lambda_{i+1})$; see Definition 1.4 and Lemma 1.5 for the definition of $d_\alpha(\beta, \gamma)$ and some of its properties. The following is true:

(3.1) The effect of reversing the orientation of λ_i is to change e_i to $-e_i$, e_{i+1} to $e_{i+1}+\pi i$, and e_{i-1} to $e_{i-1} + \pi i$.

Figure 3.1 gives rise to the two right-angled hexagons drawn in Figure 3.2. (Figure 3.2a may be inaccurate for the following reason. It is not clear whether the head of O_1 should be placed in front of the tail of O or behind the tail of O. A similar statement holds for the tail of O_1 and for Figure 3.2b.) Assume that in Figure 3.2a, λ_1 corresponds to $B_{(0;\infty)}$ and the edges are cyclically ordered counterclockwise. Then $e_6 = D$, $e_1 = L$, $e_2 = -D$. We now show that if e_5 has value c, then e_3 has value $c + \pi i$. Observe that there is an order-2 rotation τ of $\mathbf{H^3}$ about an axis orthogonal to $B_{(0;\infty)}$ which reverses the orientation on $B_{(0;\infty)}$ and takes the oriented orthocurve O to the oriented orthocurve $f(O)$. Since distance$(W(B_{(0;\infty)}), B_{(0;\infty)})$ = distance$(fW(B_{(0;\infty)}), B_{(0;\infty)})$ it follows that $\tau(fW(B_{(0;\infty)})) = -W(B_{(0;\infty)})$ and $\tau(W(B_{(0;\infty)})) = -fW(B_{(0;\infty)})$ where the $-$ sign indicates that the orientation has been reversed. This in turn implies that $\tau(O_1) = -O_1$ and therefore using (3.1) that $e_3 = e_5 + \pi i$.

Figure 3.2: Two right-angled hexagons arising from Figure 3.1.

Figure 3.3: Some ortholines (arising from Figure 3.2) coming in to $B_{(0;\infty)}$ and the distances along $B_{(0;\infty)}$ between the indicated ortholines. Each O_i^* denotes a certain $g \in G$ translate of the orthocurve O_i. Thus, two arrow-in O_i's differ by pL for some integer p.

Let ϕ be the isometry of $\mathbf{H^3}$ which is an $r + i(\pi + a)$ translation of $B_{(0;\infty)}$. Thus ϕ commutes with f, and $\phi(B_{(0;\infty)}) = B_{(0;\infty)}$, $\phi(O) = -w(O)$, $\phi(f(O)) = -fw(O)$, $\phi(W(B_{(0;\infty)})) = w(B_{(0;\infty)})$ and $\phi(fW(B_{(0;\infty)})) = fw(B_{(0;\infty)})$ which in turn implies that $\phi(O_1) = -O_2$. If the hexagon of Figure 3.2b with edges $\lambda'_1, \cdots, \lambda'_6$ is counterclockwise cyclically oriented so that λ'_1 denotes the oriented geodesic $B_{(0;\infty)}$, then again using (3.1) it follows that $e'_5 = c$ and $e'_3 = c + \pi i$.

Via elements of G, we translate each of $W(B_{(0;\infty)})$, $fW(B_{(0;\infty)})$, $w(B_{(0;\infty)})$, $fw(B_{(0;\infty)})$ to $B_{(0;\infty)}$, and after translation we obtain from Figure 3.2 the various distance relations $\{e_1, e_3, e_5, e'_1, e'_3, e'_5\}$ schematically indicated on Figure 3.3. Each O_i^* is a $g \in G$ translation of O_i such that O_i^* has an endpoint on $B_{(0;\infty)}$. There are two classes of such translates, one where O_i^* points into $B_{(0;\infty)}$ and one where O_i^* points out. Call the former (resp. latter) a pointing in (resp. out) O_i^*. Actually Figure 3.3 includes 3 more relations. Because the oriented O_1 is a G-translate of the oriented O_2 and f is a primitive element of G which fixes $B_{(0;\infty)}$, it follows that

$$\text{distance}((\text{pointing in } O_1^*), (\text{pointing in } O_2^*)) = 0 \pmod{L}$$

and

$$\text{distance}((\text{pointing out } O_1^*), (\text{pointing out } O_2^*)) = 0 \pmod{L}.$$

Finally $\text{distance}(w(O), O) = R$.

We therefore obtain the following two equations

(3.2) $c - R + L + c + \pi i = 0 \pmod{L}$ $\quad$ $c - L + R + c + \pi i = 0 \pmod{L}$

and hence

(3.3) $2R = 0 \pmod{L}$ $\quad$ $4c = 0 \pmod{L}$.

Using the L' and R' ranges for X_0 provided in Proposition 1.28, it is easy to compute that for each element of $\mathcal{R} = \exp^{-1}(X_0)$, $|\text{Re}(R)| < |\text{Re}(L)/2|$, and that in fact $R = \exp^{-1}(R')$ is close to 0 (thereby eliminating the possible solution $R = \pi i$). It then follows that $R = 0$ for $(L, D, R) \in \mathcal{R} \cap \mathcal{T}$. $\blacksquare$

Lemma 3.8: If $(L, D, R) \in \mathcal{R} \cap \mathcal{T}$, then $L = D$.

Proof: We will now use the exponential coordinates $l = \exp(L) = L'$, $d = \exp(D) = D'$. These l, d should not be confused with the $l + it, d + ib$ used above. In the following calculations Mathematica [Math] was used to perform matrix multiplication of 2×2 matrices with coefficients rational functions in the variables $\sqrt{l}$, $\sqrt{d}$. Plugging in $R = 0$ in Lemma 1.24

we get the following $\mathbf{SL}(2,\mathbf{C})$ representatives of f, F, w, W (here $\{G, f, w\}$ is the marked group corresponding to (L, D, R)). Because the R term drops out, we can express the matrices of w and W as functions of d alone.

$$f(l) = \begin{pmatrix} \sqrt{l} & 0 \\ 0 & \sqrt{l} \end{pmatrix}$$

$$F(l) = \begin{pmatrix} 1/\sqrt{l} & 0 \\ 0 & 1/\sqrt{l} \end{pmatrix}$$

$$w(d) = \begin{pmatrix} (\sqrt{d}+1/\sqrt{d})/2 & (\sqrt{d}-1/\sqrt{d})/2 \\ (\sqrt{d}-1/\sqrt{d})/2 & (\sqrt{d}+1/\sqrt{d})/2 \end{pmatrix}$$

$$W(d) = \begin{pmatrix} (\sqrt{d}+1/\sqrt{d})/2 & (-\sqrt{d}+1/\sqrt{d})/2 \\ (-\sqrt{d}+1/\sqrt{d})/2 & (\sqrt{d}+1/\sqrt{d})/2 \end{pmatrix}$$

Let Y be the relator $FwfwfWfwfw$, which is a cyclic permutation of relator i) of Lemma 3.5. Multiplying this product of 10 matrices we obtain the following matrix entries for $Y(l,d) = Y$ which we know is the identity in $\mathbf{PSL}(2,\mathbf{C})$ and hence $Y = \pm I$. (At the end of the proof of Lemma 3.9, we shall see that $Y = I$.)

$$\begin{gathered} Y_{11} = ((1+d)*(1-2*d^2+d^4+8*d*l-16*d^2*l \\ +8*d^3*l-2*l^2+4*d*l^2-4*d^2*l^2+4*d^3*l^2- \\ 2*d^4*l^2+l^4+4*d*l^4+6*d^2*l^4+4*d^3*l^4 \\ +d^4*l^4))/(32*d^{(5/2)}*l^{(5/2)}), \end{gathered}$$

$$\begin{gathered} Y_{12} = ((-1+d)*(1+4*d+6*d^2+4*d^3+d^4+4*d*l+ \\ 8*d^2*l+4*d^3*l-2*l^2-12*d^2*l^2-2*d^4*l^2+ \\ 4*d*l^3+8*d^2*l^3+4*d^3*l^3+l^4+4*d*l^4+ \\ 6*d^2*l^4+4*d^3*l^4+d^4*l^4))/(32*d^{(5/2)}*l^{(5/2)}), \end{gathered}$$

$$\begin{gathered} Y_{21} = ((-1+d)*(1+4*d+6*d^2+4*d^3+d^4+4*d*l+ \\ 8*d^2*l+4*d^3*l-2*l^2-12*d^2*l^2-2*d^4*l^2+ \\ 4*d*l^3+8*d^2*l^3+4*d^3*l^3+l^4+4*d*l^4+ \\ 6*d^2*l^4+4*d^3*l^4+d^4*l^4))/(32*d^{(5/2)}*l^{(3/2)}), \end{gathered}$$

$$\begin{gathered} Y_{22} = ((1+d)*(1+4*d+6*d^2+4*d^3+d^4-2*l^2+4*d*l^2- \\ 4*d^2*l^2+4*d^3*l^2-2*d^4*l^2+8*d*l^3- \\ 16*d^2*l^3+8*d^3*l^3+l^4-2*d^2*l^4+d^4*l^4)) \\ /(32*d^{(5/2)}*l^{(3/2)}) \end{gathered}$$

Since G is generated by f, an L translation along $B_{(0;\infty)}$, and w, a D translation along $B_{(-1;1)}$, it follows from Lemma 3.6 that the relation $Y = I$ holds with l and d switched. Thus $0 = Y_{12}$, and $0 = Y_{12}$ (with l, d switched) which implies that

$$\begin{gathered} 0 = (1 + 4*d + 6*d^2 + 4*d^3 + d^4 + 4*d*l + 8*d^2*l+ \\ 4*d^3*l - 2*l^2 - 12*d^2*l^2 - 2*d^4*l^2+ \\ 4*d*l^3 + 8*d^2*l^3 + 4*d^3*l^3 + l^4 + 4*d*l^4+ \\ 6*d^2*l^4 + 4*d^3*l^4 + d^4*l^4)- \\ (1 + 4*l + 6*l^2 + 4*l^3 + l^4 + 4*l*d + 8*l^2*d+ \\ 4*l^3*d - 2*d^2 - 12*l^2*d^2 - 2*l^4*d^2+ \\ 4*l*d^3 + 8*l^2*d^3 + 4*l^3*d^3 + d^4 + 4*l*d^4+ \\ 6*l^2*d^4 + 4*l^3*d^4 + l^4*d^4) = \\ \\ 4*(1+d)^2*(1+l)^2*(-d+l)*(-1+d*l). \end{gathered}$$

This implies that $d = l$ and hence $D = L$, or we obtain one of the following solutions which contradicts the conditions $\mathrm{Re}(L) > 0$, $\mathrm{Re}(D) > 0$. The solution $d = -1$ implies $D = \ln(d) = \ln(-1) = \pi i$. The solution $l = -1$ implies $L = \pi i$. The solution $d * l = 1$ implies that $D = -L$. ∎

Lemma 3.9: If $(L, D, R) \in \mathcal{R} \cap \mathcal{T}$, then $d = \exp(D)$ is a root of the polynomial

$$1 + 2*d + 6*d^2 + 2*d^3 + d^4.$$

Proof: The equation $Y_{12} = 0$ yields

$$\begin{gathered} 0 = 1 + 4*d + 6*d^2 + 4*d^3 + d^4 + 4*d*l + 8*d^2*l + 4*d^3*l - 2*l^2- \\ 12*d^2*l^2 - 2*d^4*l^2 + 4*d*l^3 + 8*d^2*l^3 + 4*d^3*l^3 + l^4 + 4*d*l^4+ \\ 6*d^2*l^4 + 4*d^3*l^4 + d^4*l^4. \end{gathered}$$

Setting $l = d$ we obtain

$$\begin{gathered} 0 = \\ 1 + 4*d + 8*d^2 + 12*d^3 - 2*d^4 + 12*d^5 + 8*d^6 + 4*d^7 + d^8 = \\ (1 + 2*d - 2*d^2 + 2*d^3 + d^4)*(1 + 2*d + 6*d^2 + 2*d^3 + d^4). \end{gathered}$$

On the other hand setting $l = d$ in the equation $Y_{11} = 1$ we obtain

$$32d^5 = (1+d)(1 + 4*d^2 - 12*d^3 + 6*d^4 + 8*d^5 + 4*d^6 + 4*d^7 + d^8)$$

which can be rewritten as

$$0 = (-1 + d) * (1 + 2 * d + 6 * d^2 + 2 * d^3 + d^4) * (-1 + 4 * d^3 + d^4)$$

The only solutions to these equations with $(\ln(d), \ln(d), 0) \in \mathcal{R} \cap \mathcal{T}$ are the roots of the equation

$$(1 + 2 * d + 6 * d^2 + 2 * d^3 + d^4) \tag{3.4}$$

The equation resulting from $Y_{11} = -1$ would be

$$-32d^5 = (1 + d)(1 + 4 * d^2 - 12 * d^3 + 6 * d^4 + 8 * d^5 + 4 * d^6 + 4 * d^7 + d^8).$$

But it is easy enough to check that there are no solutions d with $(\ln(d), \ln(d), 0) \in \mathcal{R} \cap \mathcal{T}$.
∎

Remarks 3.10: i) From relator i) of Lemma 3.5 and from $R = 0$ we deduced that $Y = I$ and that l and d can be switched in Y. Relator ii) was used in the proof that $R = 0$.

ii) Our matrix representatives given in the proof of Lemma 3.8 define a lift to $\mathbf{SL}(2, \mathbf{C})$ of our representation of $\pi_1(\text{Vol3})$ into $\mathbf{PSL}(2, \mathbf{C})$. Indeed we showed above that the relator i) corresponds to the identity in $\mathbf{SL}(2, \mathbf{C})$ and having solved $d = l = \omega$ and $R = 0$, a direct calculation shows that relator ii) also corresponds to the identity in $\mathbf{SL}(2, \mathbf{C})$.

Lemma 3.11: The roots of $1 + 2 * d + 6 * d^2 + 2 * d^3 + d^4$ are

$$(-1 - i * \sqrt{3})/2 - (-6 + 2 * i * \sqrt{3})^{(1/2)}/2$$

$$(-1 - i * \sqrt{3})/2 + (-6 + 2 * i * \sqrt{3})^{(1/2)}/2$$

$$(-1 + i * \sqrt{3})/2 - (-6 - 2 * i * \sqrt{3})^{(1/2)}/2$$

$$(-1 + i * \sqrt{3})/2 + (-6 - 2 * i * \sqrt{3})^{(1/2)}/2$$

∎

Remarks 3.12: i) If x is a root of $1 + 2 * d + 6 * d^2 + 2 * d^3 + d^4$, then so are $\bar{x}$, $1/x$ and $1/\bar{x}$.

ii) $-6 + 2 * i * \sqrt{3} = 4 * \sqrt{3} \exp(5 * \pi * i/6)$ and hence $\pm\sqrt{-6 + 2 * i * \sqrt{3}} = \pm 2\sqrt[4]{3} \exp(5 * \pi * i/12) = \pm(2\sqrt[4]{3})(\sqrt{2}/4)((\sqrt{3} - 1) + i(\sqrt{3} + 1))$

Lemma 3.13: If $(L, D, R) \in \mathcal{R} \cap \mathcal{T}$, then

$$\begin{aligned} D = L &= \ln((-1 - i * \sqrt{3})/2 - (-6 + 2 * i * \sqrt{3})^{(1/2)}/2) \\ &= \omega \approx 0.83144294552931 - 1.945530759503636 * i \end{aligned}$$

Proof: The other 3 solutions to (3.4), $-\omega$, $\bar{\omega}$, and $-\bar{\omega}$, all lie outside of $\mathcal{R}$. The solution ω lies in $\mathcal{R}$. ∎

Remark 3.14: In our language, Hodgson and Weeks [HW1] knew that $\pi_1(\text{Vol3})$ was generated by f, w with $D = L = \omega$ and $R = 0$. Also that the various solutions of (3.4) corresponded to symmetries of Vol3.

Corollary 3.15: Vol3 is the unique hyperbolic 3-manifold with associated parameter values in $\mathcal{S} \cap X_0$. ∎

Proof of Proposition 3.1: The previous lemmas established the first two sentences of Proposition 3.1. By Proposition 2.8 if $\text{maxcorona}(\delta) \geq \ln(3)/2$, then the parameter for a 2-generator subgroup G of $\pi_1(N)$ lies in $\mathcal{S} \cap X_0$. Since $G = \pi_1(\text{Vol3})$, N is covered by Vol3. By Jones-Reid (see [JR]) Vol3 only covers Vol3. Therefore $N = \text{Vol3}$. See Remark 3.20. ∎

Definition 3.16: Let δ be a geodesic in the hyperbolic 3-manifold N and $\{\delta_i\}_{i \geq 0}$ be the preimages of δ in $\mathbf{H^3}$. Define an equivalence relation on $\{\delta_i\}_{i \geq 1}$ by saying that δ_i and δ_j are equivalent if either

i) there exists a $g \in \pi_1(N)$ such that $g(\delta_0) = \delta_0$ and $g(\delta_i) = \delta_j$

or

ii) there exists $g \in \pi_1(N)$ such that $g(\delta_i) = \delta_0$ and $g(\delta_0) = \delta_j$.

Call an equivalence class an *orthoclass*. Counting with multiplicity, consider the collection $\mathcal{O}$ of complex numbers $\{\text{distance}(\delta_0, \delta_i) \mid i > 0$ and only one δ_i is represented in each equivalence class $\}$. Now order $\mathcal{O}$ to obtain the *ortholength spectrum of* δ, $\{O(1), O(2), \cdots\}$ where $i \leq j$ implies $\text{Re}(O(i)) \leq \text{Re}(O(j))$. Let $\mathcal{O}(i)$ denote the equivalence class corresponding to $O(i)$.

The *based ortholength spectrum of* δ consists of all distinct pairs of complex numbers of the form $\{(\text{distance}(B_{(-1;1)}, \text{ortholine from } B_{(0;\infty)} \text{ to } g(B_{(0;\infty)})), \text{distance}(B_{(0;\infty)}, g(B_{(0;\infty)}))) \mid g \in \pi_1(N), g \neq f^n\}$ Up to isometry of $\mathbf{H}^3$, the based ortholength spectrum gives the complete description of how $\{\delta_i\}$ is embedded in $\mathbf{H}^3$.

Proof of Proposition 3.2: Let δ be a shortest geodesic in Vol3. Let $\{\delta_i\}$ denote the preimages of δ in $\mathbf{H}^3$ with δ_0 identified with $B_{(0;\infty)}$ and δ_1, a nearest translate, identified with $w(\delta_0)$. By the proofs of Lemma 1.13 and Proposition 3.1 we can assume that the associated parameters satisfy $L = D = \omega$ and $R = 0$. Define a $(\pi_1(\text{Vol3}), \partial\delta_i)$ insulator family which is a hybrid of Dirichlet and Corona insulators as follows. If $\delta_i \in \mathcal{O}(1)$, then let λ_{0i} be the Dirichlet insulator between δ_0 and δ_i (see Definition 2.4). Also use the Dirichlet insulator λ_{0i}, if $\delta_i \in \mathcal{O}(k)$ and the $\text{visualangle}_{\delta_0}(\lambda_{0i}) < 113.16$. (By [G], it is less than 113.16 when $\text{Re}(O(2)) > 2\text{Arccosh}(1/\sin(113.16/2))$. Otherwise use a *Corona* insulator, i.e. $\lambda_{0i} = \kappa_{0i}$, where κ_{0i} is constructed in Proposition 2.3. Extend equivariantly to obtain the family $\{\lambda_{ij}\}$ which by construction satisfies conditions A.1. In Lemma 3.17 we will show that if $i > 0$ and $\delta_i \notin \mathcal{O}(1)$, then, $\text{visualangle}_{\delta_0}(\lambda_{0i}) < 113.16$. To complete the proof of Proposition 3.2 we now show that the insulator family $\{\lambda_{ij}\}$ is noncoalescable.

We begin by analyzing $\mathcal{O}(1)$ insulators. The geodesic plane E^* midway between $B_{(0;\infty)}$ and $w^{-1}(B_{(0;\infty)})$ intersects the $B_{(0;\infty)}$-$B_{(-1;1)}$ plane in a geodesic E at distance $\text{Re}(\omega)/2 = \text{Relength}(f)/2$ from $B_{(0;\infty)}$. Conjugate $\pi_1(N)$ via a rotation and homothety fixing $B_{(0;\infty)}$ so that one endpoint of E is at (1,0) and the other endpoint is at $(x, 0)$, where $x > 1$. Because $\text{Redistance}(E, B_{(0;\infty)}) = \text{Re}(\omega)/2$, formula (7.23.1) of [Beardon] implies $x <$

23.815. Thus the Dirichlet insulator $\lambda = \partial \bar{E}^*$ (resp. $w(\lambda)$) between $B_{(0;\infty)},\ w^{-1}(B_{(0;\infty)})$ (resp. $B_{(0;\infty)},\ w(B_{(0;\infty)})$) is symmetric about the x-axis and lies within the circle passing through (1,0), (23.815,0) (resp. (-1,0), (-23.815, 0)). By Lemma 4.7 of [G] this circle takes up a visual angle of less than 133.68 degrees. Now f is the composition of an $\exp(\mathrm{Re}(\omega))$ homothety centered about the origin and an $\mathrm{Im}(\omega)$ radian ≈ -111.4707 degree rotation. Because $\exp(4 * \mathrm{Re}(\omega)) > 27.82 > 23.815$ it follows that the plane E^* is taken "beyond" E^* by w^4. Thus, $\lambda \cap (f^n(\lambda) \cup f^n w(\lambda)) = \emptyset$ if $|n| \geq 4$. Therefore if there was a trilinking among three insulators associated to $\mathcal{O}(1)$, the orthoclass of $w(B_{(0;\infty)})$, there would be a tri-linking involving 3 circles from the collection $\{f^n w(\lambda), f^n(\lambda) \mid -3 \leq n \leq 0\}$, one of which is λ.

Because the absolute value of the rotational effect of $f^{\pm 1}$ is $|\mathrm{Im}(\omega)|$ radians, which is less than 111.48 degrees, and because λ takes up less than 133.68 degrees $B_{(0;\infty)}$-visual angle, it follows that $f^{-1}(\lambda) \cup \lambda$ takes up less than $133.68 + 111.48 = 245.06$ degrees $B_{(0;\infty)}$-visual angle. Similar arguments show that if λ and one of $f^n(\lambda)$ or $f^m w(\lambda)$ nontrivially intersect, then the union cannot take up more $B_{(0;\infty)}$-visual angle, in fact except for $f^{\pm}(\lambda)$, it takes up less. See Figure 3.4 which shows the union of $\{f^n w(\lambda), f^n(\lambda) \mid -3 \leq n \leq 0\}$. Therefore, if the union of three such circles was connected, they would take up at most $133.68 + 2(111.48) = 356.64$ degrees of $B_{(0;\infty)}$-visual angle and hence would not create a trilinking. Thus, $\mathcal{O}(1)$ cannot by itself create a tri-linking.

Figure 3.4: There is no tri-linking from $O(1)$.

Now assume Lemma 3.17. There remain three cases to consider: trilinking involving no $\mathcal{O}(1)$ insulators; trilinking involving exactly one $\mathcal{O}(1)$ insulator; trilinking involving exactly two $\mathcal{O}(1)$ insulators. The case of no $\mathcal{O}(1)$ insulators cannot occur because $3(113.16) < 360$. The case of exactly one $\mathcal{O}(1)$ insulator cannot occur because the union of three such insulators will take up less than $133.68 + 2(113.16) = 360$ degrees of visual angle. Finally, the case of exactly two $\mathcal{O}(1)$ insulators cannot occur, because the two $\mathcal{O}(1)$ insulators would take up at most 245.06 degrees of visual angle. ∎

Lemma 3.17: If $\delta_i \in \mathcal{O}(k),\ k > 1$, then $\mathrm{visualangle}_{\delta_0}(\lambda_{0i}) < 113.16$.

Proof: Step 1: $\mathrm{Re}(O(2)) > \mathrm{Re}(O(1))$

Proof of Step 1. If $\mathrm{Re}(O(2)) = \mathrm{Re}(O(1))$, then $\mathcal{T}$ contains a parameter with $L = \omega$ and $D = O(2)$. By Proposition 1.28 this can only happen if the parameter lies in X_0 and by Proposition 3.1 this can only happen if $O(2) = \omega = O(1)$.

If $v \in \pi_1(\mathrm{Vol3})$ is an element with $\mathrm{distance}(v(B_{(0;\infty)}), B_{(0;\infty)}) = O(2) = \omega$ as above, then the group generated by v, f is conjugate (by an orientation-preserving isometry taking $B_{(0;\infty)}$ to itself) to the group generated by w, f. This implies that the $\pi_1(\mathrm{Vol3})$ translates of $B_{(0;\infty)}$ are symmetrically placed about $B_{(0;\infty)}$. We now restrict our focus to $\delta_i \notin \mathcal{O}(1)$ with $\mathrm{distance}(\delta_i, B_{(0;\infty)}) = \omega$, and define K to be the unoriented orthocurve between $v(B_{(0;\infty)})$ and $B_{(0;\infty)}$. There are two cases to consider.

The first case is that K hits $B_{(0;\infty)}$ at $(0,0,1)$. Now (after possibly rechoosing δ_i, and using the fact that $R = 0$) we can assume that the angle between K and K' is π/m, where K' is the orthocurve between $B_{(0;\infty)}$ and $w(B_{(0;\infty)})$ and $m > 1$ is a maximal integer. Using the hyperbolic law of cosines ([F], page 83), it follows that if $m=2$, then there exists an $\bar\omega$ ortholength. As in the first paragraph of this proof, we obtain a contradiction. See Figure 3.5. If $m>2$, then the cosine law shows that there exists a real ortholength less than $\mathrm{Re}(\omega)$.

Figure 3.5: $\cosh(P)=\cosh^2(\omega)+\sinh^2(\omega)\cosh(i(\pi/2+\pi))$ has solution $P=\bar\omega$.

The second case is that $\mathrm{Redistance}(K,K')=x>0$ where x is minimal among all possible choices. Then by symmetry we can assume that there is an m with $mx=\mathrm{Re}(\omega)=\mathrm{Re}(L)$. Indeed δ_i can be chosen so that if δ_i is obtained from $w(B_{(0;\infty)})$ by an $x+iy$ translation τ along $B_{(0;\infty)}$, then $m(x+iy)=\omega$. (This uses the fact that the R associated to any orthoclass $\mathcal{O}(i)$ with $O(i)=\omega$ is equal to 0.) Therefore an $x+iy$ translation τ along $B_{(0;\infty)}$ descends to a free $\mathbf{Z}/m\mathbf{Z}$ action ϕ on Vol3. This is free, because any lift of ϕ^n is a conjugate (by an orientation-preserving isometry) of τ^n which is fixed-point free or the identity. This contradicts the fact that Vol3 only covers Vol3. ∎

Step 2: Completion of Proof of Lemma 3.17.

Proof of Step 2. We obtain this result with computer assistance in a manner similar to the proofs of Propositions 1.28 and 2.8. Our parameter space $\mathcal{W}$ is the usual initial box. As before, a parameter gives rise to a marked group $\{G,f,v\}$. (Here, f is as before the standard generator of $\pi_1(\mathrm{Vol3})$ fixing $B_{(0;\infty)}$.) We consider the parameters U such that f,v generate a torsion-free, parabolic-free group G where f is of minimal length, $\mathrm{Re}(D)>\mathrm{Re}(\omega)$, and $\mathrm{Corona}(D)\geq 113.16$. (It is easily checked that $U\subset\mathcal{W}$, and it is interesting to note that U contains parameters with $\mathrm{Re}(D)$ almost 1.24.) We will partition $\mathcal{W}$ into regions $\mathcal{W}_i$ such that each $\mathcal{W}_i$ can be eliminated for one of the following reasons.

y) $\mathcal{W}_i\notin U$

z) $\mathcal{W}_i=X_0$.

a) There exists no $\beta\in\mathcal{W}_i$ such that $\mathrm{length}(f_\beta)\in\mathcal{L}(X_0)$. In particular, $\mathrm{length}(f_\beta)\neq\omega$ throughout $\mathcal{W}_i$.

b) $\mathcal{W}_i$ has some L values in $\mathcal{L}(X_0)$ but there exists no $\beta\in\mathcal{W}_i$ such that the real part of $\mathrm{distance}(v_\beta(B_{(0;\infty)}),B_{(0;\infty)})$ is greater than the minimum d value for X_0. In particular, $\mathrm{Re}(\mathrm{distance}(v_\beta(B_{(0;\infty)}),B_{(0;\infty)}))\leq\mathrm{Re}(\omega)$ throughout $\mathcal{W}_i$.

c) There exists a killerword h in f,v,f^{-1},v^{-1}, such that $h_\beta\neq\mathrm{id}$ and $\mathrm{Relength}(h_\beta)<\mathrm{Relength}(f_\beta)$ for all $\beta\in\mathcal{W}_i$.

There are files containing the partition of $\mathcal{W}$ and the associated conditions/killerwords, and the program *corona* verifies that they indeed work. As noted earlier, *corona* works with the exponentiated versions of the above conditions. The computer methods (involving the Corona function) needed to complete the proofs in Chapters 2 and 3 are sufficiently similar that it was natural to incorporate both proofs into one partitioning of $\mathcal{W}$, one list of associated conditions/killerwords, and one computer program (*corona*).

Lemma 3.17 follows from the fact that each box can be eliminated. Indeed if $\delta_i\in\mathcal{O}(k),\ k>1$, then Step 1 implies that $\mathrm{Re}(O(k))>\mathrm{Re}(O(1))=\mathrm{Re}(\omega)$. If λ_{0i} is a Dirichlet insulator, then by construction either $\delta_i\in\mathcal{O}(1)$ or $\mathrm{visualangle}_{\delta_0}(\lambda_{0i})<113.16$. If λ_{0i} is a Corona insulator and $\mathrm{visualangle}_{\delta_0}(\lambda_{0i})\geq 113.16$, then the group generated by f and v

gives rise to a parameter in U where $v \in \pi_1(\text{Vol3})$ is an element taking δ_0 to δ_i. The above program rules out this possibility. ∎

Here is an experimental "proof" of Lemma 3.17. In [HW2] an algorithm is given to compute, with multiplicities, the length spectrum of a hyperbolic 3-manifold M, given a Dirichlet domain for M. Weeks has observed (see [W2]) that a very similar argument gives an algorithm to compute the based ortholength spectrum. In fact an analogue to Proposition 1.6.2 in [HW2],(with an analogous proof) is the following.

Lemma 3.18 (Weeks): Let M be a closed orientable 3-manifold having a Dirichlet domain $\mathcal{D}$ with basepoint x and with spine radius r. Let δ be a geodesic of length $l + it$. To compute all the based ortholengths of real length less than or equal to λ with basing less than or equal to $l/2$ from some point on δ_0 (a preimage of δ) it suffices to find all translates $g\mathcal{D}$ satisfying $\rho(x, gx) \leq 2r + 2\text{Arccosh}(\cosh(l/2)\cosh(\lambda/2))$.

Proof: As in [HW2] we can assume that $\rho(\delta_0, x) \leq r$. The 0-basing on δ_0 will be given by the oriented perpendicular P from δ_0 to x. Figure 3.6 shows that if there is a translate $\delta_i = g(\delta_0)$ based at distance $\leq l/2$, at real distance $\leq \lambda$ from δ_0, then $\rho(g(x), x) \leq \lambda + l + 2r$. As in [HW2] an application of the hyperbolic cosine law yields the better estimate of Lemma 3.18. ∎

Figure 3.6: Controlling the based ortholength spectrum.

Provided one is given a Dirichlet domain, [HW2] gives an efficient algorithm to find these g's. Finally to each such g one computes the basing and distance from $g(\delta_0)$ to δ_0.

The collection of ortholengths with basings $\leq l/2$ contains at least 2 representatives for each ortholength class, thus the Weeks algorithm can be used to give lower bounds on the various $O(i)$'s. *SnapPea* (see [W1]) computes a Dirichlet domain for Vol3 with spine radius ≤ 0.68. Because $l < 0.83145$, taking $\lambda = 1.24$ we obtain $2r + 2\text{Arccosh}(\cosh(l/2)\cosh(\lambda/2)) < 2.89$. This algorithm has been implemented on an undistributed version of *SnapPea*, and provided the following estimates. Note that $\text{Re}(O(2)) \geq 1.24$ is sufficient to guarantee that $\mathcal{C}(O(2)) < 113.16$ degrees.

$$O(1) \approx .83144 - 1.94553i,$$

$$O(2) \approx 1.3170 - \pi i,$$

$$O(3) = O(4) \approx 1.4197 + 1.0963i,$$

$$O(5) \approx 1.9769 - 1.2995i.$$

These estimates were found using a "tiling" of radius $3.00 > 2.89$, which is sufficient for "proving" that for Vol3, $\text{Re}(O(2)) > 1.24$. ∎

Remark 3.19: This experimental proof should be easy to make rigorous by implementing Weeks's algorithm using exact arithmetic, say via the program *Snap* (see [CGHN]).

Remark 3.20: Similar technology can probably be used to show that Vol3 only covers Vol3. Because Vol3 is non-Haken, and non-orientable 3-manifolds are Haken, Vol3 can only cover an orientable 3-manifold. By Proposition 1.28 and the first sentence of Proposition 3.1, it follows that either a shortest geodesic δ in M has tuberadius$(\delta) = \text{Re}(\omega)/2$ and length$(\delta) = \omega$, or δ has a $\ln(3)/2$ tube. The latter implies that some geodesic β in Vol3 has a $\ln(3)/2$ tube. Elementary volume considerations together with length spectrum data provided by *SnapPea* imply that length$(\beta) < 1$ and hence is one of 8 geodesics. Tuberadius data provided by an unreleased version of *SnapPea* assert that such geodesics have tube radius smaller than 0.43. Thus, the length of δ is ω and so volume$(M) > \pi \sinh(\text{Re}(\omega)/2))^2 * \text{Re}(\omega) > 0.476 >$ volume(Vol3)/3. This implies that there exists a free, orientation-preserving involution τ of Vol3 such that $\tau(\delta) \neq \delta$. By length spectrum data from *SnapPea*, there are only two distinct geodesics with the same length as δ. The above analysis shows that these correspond to $B_{(0;\infty)}$ and $B_{(-1;1)}$, which non-trivially intersect. Thus, no such involution exists. Using the exact arithmetic of *Snap* (see [CGHN]), one should be able to make rigorous these *SnapPea* calculations.

Chapter 4: Applications

We provide some applications. First, we give a partial answer to Problem 1.16. Second a lower bound for the volume of hyperbolic 3-manifolds is produced. Finally, we give a relationship between isotopic closed curves in a hyperbolic 3-manifold and essential links in B^3.

Theorem 4.1: If δ is a shortest geodesic in the closed orientable hyperbolic 3-manifold N, then either

i) tuberadius$(\delta) > \ln(3)/2$, or
ii) $1.0953/2 >$ tuberadius$(\delta) > 1.0591/2$ and length$(\delta) > 1.0595$, or
iii) tuberadius$(\delta) = 0.8314\ldots/2$ and $N =$ Vol3.

Proof: If $0.5 <$ tuberadius$(\delta) = d/2 \leq \ln(3)/2$, then by Corollary 1.29 and Table 1.2 we see that $1.0953/2 >$ Tuberadius$(\delta) > 1.0591/2$ and length$(\delta) > 1.0595$. If tuberadius$(\delta) = d/2 \leq 0.5$ then the parameter $(L, D, R) \in \mathcal{P}$ associated to N must be in $\mathcal{R} \cap \mathcal{T} = \exp^{-1}(X_0 \cap \mathcal{S})$. It then follows by Corollary 3.15 that $N =$ Vol3. ∎

Remark 4.2: The previous best lower bound for the volume of hyperbolic 3-manifolds was on the order of 0.001 (see [GM1]). Using the results of the present paper and the method of [M1] it is easy to improve this to 0.1. However, Gehring and Martin provide an improved tube-volume formula in [GM2] and we use their formula to get a lower bound of $0.16668\ldots$. The Gehring-Martin tube-volume formula for manifolds (as opposed to orbifolds) is

$$\mathcal{V}(t) = \sqrt{3}\tanh(t)\cosh(2t)\text{Arcsinh}^2(\sinh(t)/\cosh(2t))$$

where t is the radius of the embedded solid tube. Note that the length of the core geodesic is irrelevant.

Corollary 4.3: $\frac{5}{2\sqrt{3}}\text{Arcsinh}^2(\frac{\sqrt{3}}{5}) = 0.16668\ldots$ is a lower bound for the volume of closed hyperbolic 3-manifolds.

Proof: Corollary 1.7 of [GM2] applies the tube-volume formula $\mathcal{V}(t)$ to tubes of radius at least $\ln(3)/2$ and produces

$$\text{Vol}(N) \geq \mathcal{V}(\ln(3)/2) = \frac{5}{2\sqrt{3}}\text{Arcsinh}^2(\frac{\sqrt{3}}{5}) = 0.16668\ldots$$

So, if a hyperbolic 3-manifold N has a shortest geodesic with tuberadius greater than or equal to $\ln(3)/2$ then we are done. If tuberadius is less than $\ln(3)/2$ then Theorem 4.1 implies that either N is Vol3, or l and d are bounded as follows: $l \geq 1.0595$ and $d \geq 1.0591$ In the first case, $\text{Vol}(N) = \text{Vol}(\text{Vol3}) = 1.01\ldots$, while in the second case, plugging the l and d bounds into the tube-volume formula $\pi l \sinh^2(d/2)$ we get $\text{Vol}(N) > 1.02$ ∎

Remark 4.4: Recently, A. Przeworski has proven a lower bound for volume of 0.276796 (see [Pr]). His method is to develop an improved version of Gehring and Martin's tube-volume formula and then to apply it to our tuberadius results.

Theorem 4.5: Let k_1, k_2 be simple closed curves in N such that k_1 is a geodesic. Then k_1 is isotopic to k_2 if and only if as B^3-links $q^{-1}(k_1)$ is equivalent to $q^{-1}(k_2)$ where $q : \mathbf{H^3} \to N$ is the universal covering projection.

Proof: Apply Corollary 5.6 of [G]. Recall that Γ and Δ are equivalent B^3-links if there is a homeomorphism of B^3 which takes Γ to Δ and fixes S^2 pointwise. ∎

Remark 4.6: A similar argument extends Theorem 4.5 to homotopy essential links which lift to trivial B^3-links. The general case of Conjecture 5.5A of [G] is still open.

Chapter 5: Conditions and Sub-Boxes

In this chapter we expand on some topics mentioned briefly in Chapter 1. As such, it would be useful to look again at Definitions 1.8, 1.12, and 1.22, and Remark 1.9, where $\mathcal{P}$, $\mathcal{T}$ and their partners (under exponention) $\mathcal{W}$, $\mathcal{S}$ are introduced, and at Definition 1.18 where the notion of a killerword is introduced (the definition is phrased in terms of $\mathcal{P}$, but the definition also makes sense for $\mathcal{W}$). Note that working with the region $\mathcal{P}$ is intuitively appealing, but working with the box $\mathcal{W}$ is vastly superior computationally (see Remark 1.21).

Theorem 0.2, hence Theorem 0.1, follows from Propositions 1.28, 2.8, and 3.2. The computational aspects of the proofs of each of these propositions are similar. We will focus on Proposition 1.28 here. The proof of Proposition 1.28 amounts to decomposing $\mathcal{W}$ into a collection of sub-boxes of two types:

1) The exceptional boxes $X_0, X_1, \ldots, X_6$

2) Sub-boxes each of which has an associated *condition* that will describe how to kill that entire sub-box, perhaps with the help of a killerword. To *kill* a sub-box means to show that $\mathcal{S} - \bigcup_{n=0,\ldots,6} X_n$ has no point in the sub-box.

The set-up for efficiently describing these sub-boxes will be given in Construction 5.3.

We now list the conditions used to kill the non-exceptional sub-boxes. There are two types of conditions: the trivial and the interesting. The trivial conditions kill sub-boxes in $\mathcal{W}$ by simply noting that the sub-box in question misses $\exp(\mathcal{P})$. The interesting conditions are where the real work is done, and they require a killerword in f, w, f^{-1}, w^{-1} to work their magic (see Remark 1.17).

To be consistent with the computer program *verify* we use the following notation. $L' = z_0 + iz_3$, $D' = z_1 + iz_4$, and $R' = z_2 + iz_5$. Here $(L', D', R') \in \mathcal{W}$ and $L' = \exp(L) = \exp(l + it)$, $D' = \exp(D) = \exp(d + ib)$, $R' = \exp(R) = \exp(r + ia)$.

The Trivial Conditions 5.1:

Condition 's' (short): Tests that all points in the sub-box have $|z_0+iz_3| < 1.10274$. This ensures that $\exp(l) = |\exp(L)| = |L'| = |z_0 + iz_3| < 1.10274 < \exp(0.0978)$, and Definition 1.12 tells us that we are outside of $\exp(\mathcal{P})$.

Condition 'l' (long): Tests that all points in the sub-box have $|z_0+iz_3| > 3.63201$. This ensures that $\exp(l) = |\exp(L)| = |L'| = |z_0 + iz_3| > 3.63201 > \exp(1.289785)$ and we are outside of $\exp(\mathcal{P})$.

Condition 'n' (near): Tests that all points in the sub-box have $|z_1 + iz_4| < 1$. This ensures that $\exp(d) = |\exp(D)| = |D'| = |z_1 + iz_4| < 1 = \exp(0)$ and we are outside of $\exp(\mathcal{P})$.

Condition 'f' (far): Tests that all points in the sub-box have $|z_1 + iz_4| > 3$. This ensures that $\exp(d) = |\exp(D)| = |D'| = |z_1 + iz_4| > 3 = \exp(\ln 3)$ and we are outside of $\exp(\mathcal{P})$.

Condition 'w' (whirle big): Tests that all points in the sub-box have $|z_2 + iz_5|^2 > |z_0 + iz_3|$ This ensures that $\exp(r) = |\exp(R)| = |R'| = |z_2 + iz_5| > \sqrt{|z_0 + iz_3|} = \sqrt{\exp(l)} = \exp(l/2)$ and we are outside of $\exp(\mathcal{P})$.

Condition 'W' (whirle small): Tests that all points in the sub-box have $|z_2+iz_5| < 1$. This ensures that $\exp(r) = |\exp(R)| = |R'| = |z_2+iz_5| < 1 = \exp(0)$ and we are outside of $\exp(\mathcal{P})$.

The Interesting Conditions 5.2:

Condition 'L': This condition comes equipped with a killerword k in f and w, and tests that all points in the sub-box have $|\exp(\text{length}(k))| < |L'| = |\exp(L)|$, where $\text{length}(k)$ means the length of the isometry determined by k. This, of course, contradicts the fact that L is the length of the shortest geodesic.

It is easy to carry out the test $|\exp(\text{length}(k))| < |L'|$ because Lemma 1.25a can be used. Note that in *verify* the function which computes exp(length) is called *length*.

Of course, Condition 'L' also checks that the isometry corresponding to the word k is not the identity.

Condition 'O': This condition comes equipped with a killerword k in f and w, and tests that all points in the sub-box have $|\exp(\text{distance}(k(B), B))| < |D'| = |\exp(D)|$. This, of course, contradicts the "nearest" condition.

It is easy to carry out the test $|\exp(\text{distance}(k(B), B))| < |D'|$ because Lemma 1.25b can be used. Note that in *verify* the function which computes $\exp(\text{distance}(k(B), B))$ is called *orthodist*.

Also, Condition 'O' checks that the isometry corresponding to the word k does not take the axis of f to itself.

Condition '2': This is just the 'L' condition without the "not-the-identity" check, but with the additional proviso that the killerword k is of the form $f^p w^q$. This ensures that k is not the identity, because for k to be the identity f and w would have to have the same axis, which contradicts the fact that d can be taken to be greater than or equal to $l/4$.

Condition 'conjugate': There is one other condition that is used to eliminate points in $\mathcal{W}$. Following Definition 1.12 (and Lemma 1.13) we eliminate all boxes with $0 < t \leq \pi$. Of course, after exponentiating $L = l + it$, this corresponds to eliminating all boxes with $z_3 > 0$. Specifically, we toss all sub-boxes of $\mathcal{W}$ whose fourth entry is a 1. This condition does not appear in *verify* and *corona* because it is applied "outside" of these programs, as described in Construction 5.3.

Construction 5.3: We now give the method for describing the roughly 930 million sub-boxes that the initial box $\mathcal{W}$ is sub-divided into.

All sub-boxes are gotten by sub-division of a previous sub-box along a real hyper-plane mid-way between parallel faces of the sub-box before sub-division. Of course, these midway planes are of the form $x_i =$ a constant. We use 0's and 1's to describe which half of a sub-divided sub-box to take (0 corresponds to lesser x_i values). For example, 0 describes the box $\mathcal{W} \cap \{(x_0, x_1, x_2, x_3, x_4, x_5) : x_0 \leq 0\}$, 010 describes the box $W \cap \{(x_0, x_1, x_2, x_3, x_4, x_5) : x_0 \leq 0,\ x_1 \geq 0,\ x_2 \leq 0\}$, and so on.

In this way, we get a 1-to-1 correspondence between strings and sub-boxes. If s is a string of 0's and 1's, then let $Z(s)$ denote the box corresponding to s. The range of values for the i-th coordinate in the sub-box $Z(s)$ is related to the binary fraction $0.s_i s_{i+6} \ldots s_{i+6k}$. The two sub-boxes gotten from subdividing $Z(s)$ are $Z(s0)$ and $Z(s1)$.

The directions of sub-division cycle among the various coordinate axes: the n-th sub-division is across the $(n \bmod 6)$-th axis. The dimensions of the top-level box $\mathcal{W}$ were chosen so that sub-division is always done across the longest dimension of the box, and so that all of the sub-boxes are similar. The dimensions of $\mathcal{W}$ have the beneficial effect of making the sub-boxes as "round" as possible, hence making the Taylor Series calculations efficient and fast. This explains the factor of $2^{(5-i)/6}$ in Definition 1.22.

To kill a sub-box $Z(s)$, the checker program has two (recursive) options: use a condition and, if necessary, an associated killerword to kill $Z(s)$ directly, or first kill $Z(s0)$ and then kill $Z(s1)$.

There is also a third option: don't kill $Z(s)$, and instead mark s as omitted. Any omitted sub-boxes are checked with another instance of the checker program, unless the sub-box is one of the 11 exceptional sub-boxes (which produce the 7 exceptional boxes after joining abutters). Note that according to the definition of "kill" given at the beginning of this Chapter, the exceptional boxes are automatically killed.

Thus, a typical output from *verify* would be

verified000000111101111111 − {0000001111011111110 00000011110111111110}

Which means that the sub-box $Z(000000111101111111)$ was killed except for its sub-boxes $Z(0000001111011111110)$ and $Z(00000011110111111110)$. The ouput

verified0000001111011111110 − { }.

and

$$\text{verified}0000001111011111111110 - \{\ \}.$$

shows that these boxes were subsequently killed as well, and thus the entire sub-box $Z(000000111101111111)$ has been killed.

Instead of immediately working on killing the top-level box, we subdivide in the six co-ordinate directions to get the 64 sub-boxes

$$Z(000000), Z(000001), Z(000010), Z(000011), \ldots, Z(111111).$$

We then throw out the ones with fourth co-ordinate equal to 1 (see condition 'conjugate'), leaving us with the 32 sub-boxes

$$Z(000000), Z(000001), Z(000010), Z(000011), \ldots, Z(111011).$$

We then use *verify* to kill these.

The choices in *verify* are made for it by a sequence of integers given as input. The sequence of integers containing the directions for killing $Z(000000)$ is contained in the file data/000000 (actually, data/000000.d). In such a sequence, 0 tells *verify* to sub-divide the present box (by $x_i = c$), to position itself on the "left-hand" box ($x_i \leq c$) created by that sub-division, and to read in the next integer in the sequence. A positive integer n tells *verify* to kill the sub-box it is positioned at directly, using the condition (and killerword, if necessary) on line n in the "conditionlist" file, and to then position itself at the "next" natural sub-box. -1 tells *verify* to omit the sub-box, and mark it as skipped (the sequence of integers used in killing the skipped box $Z(s)$ is contained in a file data/s).

The checker program *verify*, its inputs, and the list of conditions will be available from the *Annals of Math* web site. Details about how to get them can be found at

http://www.annalsofmath.edu

Similarly for the program *corona* which is used on the 7 exceptional boxes.

Example 5.4: To illustrate the checking in action, this is a (non-representative) example, which shows how the sub-box $Z(s)$ (minus a hole) is killed, where

$$s = 0010001100011101110011110001011111111011111001110011110000011110111110111.$$

The input associated with this sub-box is

$$(0, 0, 0, 1929, 12304, 0, 0, 7, 0, 1965, 0, 1929, 1929, 1996, -1),$$

which causes the program to kill $Z(s)$ in the following fashion:

kill $Z(s)$:
 kill $Z(s0)$:
  kill $Z(s00)$:
   kill $Z(s000)$ with condition 1929 = "L(FwFWFWfWFWFwFwfww)"

kill $Z(s001)$ with condition 12304 = “L(FwfWFFWFwFwfwfWfwfw)”
kill $Z(s01)$:
kill $Z(s010)$:
kill $Z(s0100)$ with condition 7 = “L(w)”
kill $Z(s0101)$:
kill $Z(s01010)$ with condition 1965 = “L(fwFwFWFFWFwFwfwww)”
kill $Z(s01011)$:
kill $Z(s010110)$ with condition 1929
kill $Z(s010111)$ with condition 1929
kill $Z(s011)$ with condition 1996 = “L(FwFwFWFWfWFWFwFwfww)”
omit $Z(s1)$
as shown in figure 5.1.

Figure 5.1: Six levels of subdivision, in two projections, with all the trimmings

$Z(s1)$ is ignored, so the checker would indicate this omission in its report. In fact, $Z(s1)$ is one of the 11 exceptional sub-boxes (7 boxes after joining abutters), specifically X_{5a}, hence killed automatically.

The use of condition “L(w)” so deep in the tree is unusual. In this case, it’s because the manifold in the exceptional sub-box has $length(f) = length(w)$, so that the program will frequently come to places where it can bound $length(f) > length(w)$ nearby.

One might wonder why the checker subdivides $Z(s01011)$, since it’s going to use the same condition to kill both halves. The reason is the error bound for $Z(s01011)$ wasn’t good enough to prove that the sub-box is killed directly. In fact, this approach is quite useful when dealing with round-off error—if a killerword barely misses killing off a sub-box because of round-off error, then recursively sub-divide the sub-box and use the same killerword on the pieces until it succeeds.

The binary numbers used by the computer require too much space to print. In the example calculation which follows, we instead use decimal representations (although we print less digits than could be gotten from the 53 binary digits used for the actual calculations).

The sub-box $Z(s01011)$ is the region where

$$\begin{pmatrix} -1.381589027741\ldots \leq Re(L') \leq -1.379848991182\ldots \\ -1.378124546093\ldots \leq Re(D') \leq -1.376574349753\ldots \\ 0.999893182771\ldots \leq Re(R') \leq 1.001274250703\ldots \\ -2.535837191243\ldots \leq Im(L') \leq -2.534606799593\ldots \\ 2.535404997792\ldots \leq Im(D') \leq -2.534308843448\ldots \\ -0.001953125000\ldots \leq Im(R') \leq 0.000000000000\ldots \end{pmatrix}$$

At this point, we would like to compute

$$f,\ w,\ g = f^{-1}wf^{-1}w^{-1}f^{-1}w^{-1}fw^{-1}f^{-1}w^{-1}f^{-1}wf^{-1}wfww,\ \text{length}(g),$$

and so on. However, these items take on values over an entire sub-box and thus are computed via AffApprox’s (first-order Taylor Approximations with remainder bounds), which are not formally defined until the next Chapter. As such, we complete Example 5.4 at the end of Chapter 6.

Remark 5.5: For those planning on looking at the program *verify* we now tie in the above description of its workings to a portion of the actual code in the program. We note that the CWeb version of *verify* is extensively documented, and is organized so that the most important details are presented first.

If the executable version of *verify* is called *verify* and we are in the correct place with respect to the location of the data, then a typical UNIX command line would be

```
zcat data/000000.gz | verify 000000 > output000000
```

This would run *verify* at the node 000000, and, when needed, would pipe in the unzipped data from `data/000000.gz`. This unzipped data contains the tree decomposition of the parameter space at the sub-box 000000. The output from *verify* would be redirected to the file `output000000`.

In *verify*, `main` would check for syntax errors in the command line, and if there were no such errors, would read the location 000000 into the character array `where` and compute that the `depth` of `where` was 6, which means that 000000 contains 6 subdivisions. It would then immediately print

```
verified 000000 -{
```

into the file `output000000`, and then call the function *verify*, as follows:

```
verify(where, depth, 0);
```

The function `verify(where, depth, autocode)` is now invoked; this time with `autocode` equal to 0. *Verify* would first check that `depth` was not too deep. Next, *verify* checks if `autocode` is equal to 0, which it is, so it reads in the next (in this case, the first) integer from the unzipped file `data/000000.gz`, and sets `code` equal to this integer. Now, *verify* recursively calls itself on the left child (0000000) of the `where` box and the right child (0000001) of the `where` box:

```
where[depth] = '0';
verify(where, depth + 1, code);
where[depth] = '1';
verify(where, depth + 1, code);
```

In general, `verify(where, depth, autocode)` does the following. It checks to see that `depth` is not too deep. Then if `autocode` is equal to 0, it recursively calls itself on its left and right children. If `autocode` is not equal to 0, then `code` is set equal to `autocode`, and 3 possibilities can occur. Either,

1) `code` is less than zero, in which case we are at an exceptional sub-box, and *verify* prints out its location (`where`) in `output000000` and recursively moves on to the next node in the tree, or

2) `code` is greater than zero and it invokes a condition/killerword from the file *conditionlist* which kills the entire sub-box `where` in which case `verify` simply recursively moves on to the next node in the tree, or

3) `code` is greater than zero and it invokes a condition/killerword from the file *conditionlist* which does not kill the entire sub-box `where`, in which case `verify` subdivides the sub-box `where` and recursively calls itself on the left child and the right child, using the same `code`:

```
where[depth] = '0';
verify(where, depth + 1, code);
where[depth] = '1';
verify(where, depth + 1, code);
```

In this way *verify* tests the entire starting box, in this case the sub-box 000000, and if successful at killing it minus the omissions which it prints out, it finishes `main` by printing out a right bracket into `output000000`.

Chapter 6: Affine Approximations

Remark 6.1: To show that a sub-box of the parameter box $\mathcal{W}$ is killed by one of the interesting conditions (plus associated killerword) we need to show that at each point in the sub-box, the killerword evaluated at that point satisfies the given condition (see Chapter 5). That is, we are simply analyzing a certain function from the sub-box to $\mathbf{C}$.

As described in Remark 6.5, this analysis can be pulled back from the sub-box in question to the unit complex 3-disc A, where $A = \{(z_0, z_1, z_2) \in \mathbf{C}^3 : |z_k| \leq 1 \text{ for } k \in \{0, 1, 2\}\}$. Loosely, we will analyze such a function on A by using Taylor series approximations consisting of an affine approximating function together with a bound on the "error" in the approximation (this could also be described as a "remainder bound").

Problems 6.2: There are two immediate problems likely to arise from this Taylor approximation approach. The first problem is the appearance of unpleasant functions such as Arccosh. We have already taken care of this problem by "exponentiating" our preliminary parameter space $\mathcal{P}$. This resulted in all functions under consideration being built up from the co-ordinate functions L', D', and R' on $\mathcal{W}$ by means of the elementary operations $+,\ -,\ \times,\ /,\ \sqrt{\cdot}$.

Second, for a given "built-up function" the computer needs to be able to compute the Taylor approximation, and the error term. This will be handled in this chapter by developing combination formulas for elementary operations (see the Propositions below). Specifically, given two Taylor approximations with error terms representing functions g and h and an elementary operation on g and h, we will show how to get the Taylor approximation with error term for the resultant function from the two original Taylor approximations.

A similar approach was developed independently by Stolfi and Figuereido (see [FS]).

Remark 6.3: We set up the Taylor approximation approach rigorously as follows in Definition 6.4. The notation will be a bit unusual, but we are motivated by a desire to stay close to the notation used in the checker computer programs, *verify* and *corona*.

However, it should be pointed out that the formulas in this Chapter will be superceded by the ones in Chapter 8, which incorporate a round-off error analysis. It is the Chapter 8 formulas that are used in *verify* and *corona*.

Definition 6.4: An *AffApprox* x is a five-tuple $(x.f;\ x.f_0,\ x.f_1,\ x.f_2;\ x.e)$, consisting of four complex numbers $x.f,\ x.f_0,\ x.f_1,\ x.f_2$ and one real number $x.e$, which represents all functions $g : A \to \mathbf{C}$ such that

$$|g(z_0, z_1, z_2) - (x.f + x.f_0 z_0 + x.f_1 z_1 + x.f_2 z_2)| \leq x.e$$

for all $(z_0, z_1, z_2) \in A$. That is, x represents all functions from A to $\mathbf{C}$ that are $x.e$-well-approximated by the affine function $x.f + x.f_0 z_0 + x.f_1 z_1 + x.f_2 z_2$. We will denote this set of functions associated with x by $S(x)$.

Remark 6.5: As mentioned in Remark 6.1, given a sub-box to analyze, instead of working with functions defined on the sub-box, we will work with corresponding functions defined on A. Specifically, rather than build up a function by elementary operations performed on the co-ordinate functions L', D', R' restricted to the given sub-box, we will perform the elementary operations on the following functions defined on A,

$$(p_0 + ip_3; s_0 + is_3, 0, 0; 0) \quad (p_1 + ip_4; 0, s_1 + is_4, 0; 0) \quad (p_2 + ip_5; 0, 0, s_2 + is_5; 0)$$

where $(p_0 + ip_3, p_1 + ip_4, p_2 + ip_5)$ is the center of the sub-box in question, and the s_i describe the six dimensions of the box. In the computer programs, these three functions are called *along*, *ortho*, and *whirle*, respectively, and p_i and s_i are denoted *pos*[i] and *size*[i], respectively.

After the following Remarks, we state and prove the combination formulas.

Remarks 6.6: i) We will break with the convention used previously in this paper and start the numbering of the Propositions with 6.1. However, we will end this Chapter with Example 6.7.

ii) The negation of a set of functions is the set consisting of the negatives of the original functions, and similarly for other operations.

iii) The propositions that follow include in their statements the definitions of the various operations on AffApprox's. What needs to be proven is that the S functions behave as expected. For example, we need to show that under the definition given for addition, the set of functions $S(x + y)$ contains all functions gotten by adding a function from $S(x)$ to a function from $S(y)$.

Proposition 6.1 (unary minus): If x is an AffApprox, then $S(-x) = -(S(x))$ where

$$-x \equiv (-x.f;\ -x.f_0,\ -x.f_1,\ -x.f_2;\ x.e).$$

Proof:

$$|g(z_0, z_1, z_2) - (x.f + x.f_0 z_0 + x.f_1 z_1 + x.f_2 z_2)| \leq e$$

if and only if

$$|-g(z_0,z_1,z_2)-(-x.f-x.f_0z_0-x.f_1z_1-x.f_2z_2)|\le e$$

∎

Proposition 6.2 (addition): If x and y are AffApprox's, then $S(x+y)\supseteq S(x)+S(y)$, where

$$x+y\equiv(x.f+y.f;x.f_0+y.f_0,x.f_1+y.f_1,x.f_2+y.f_2;x.e+y.e)$$

Proof: If $g\in S(x)$ and $h\in S(y)$ then we must show that $g+h\in S(x+y)$.

$$|(g+h)(z_0,z_1,z_2)-((x.f+y.f)+(x.f_0+y.f_0)z_0+(x.f_1+y.f_1)z_1+(x.f_2+y.f_2)z_2)|$$

$$\le|g(z_0,z_1,z_2)-(x.f+(x.f_0)z_0+(x.f_1)z_1+(x.f_2)z_2)|+$$
$$|h(z_0,z_1,z_2)-(y.f+(y.f_0)z_0+(y.f_1)z_1+(y.f_2)z_2)|$$

$$\le x.e+y.e$$

∎

Proposition 6.3 (subtraction): If x and y are AffApprox's, then $S(x-y)\supseteq S(x)-S(y)$, where

$$x-y\equiv(x.f-y.f;x.f_0-y.f_0,x.f_1-y.f_1,x.f_2-y.f_2;x.e+y.e)$$

∎

We now state variations on Propositions 6.2 and 6.3 whose usefulness will not be apparent until Chapter 8, when we incorporate round-off error into these formulas. Similarly for Propositions 6.7, 6.9, and 6.10. In what follows, a "double" corresponds to a real number, and has an associated AffApprox, with last four entries zero. When we do machine arithmetic in Chapters 7 and 8, doubles will be machine numbers.

Proposition 6.4 (addition of an AffApprox and a double): If x is an AffApprox and y is a double, then $S(x+y)\supseteq S(x)+S(y)$, where

$$x+y\equiv(x.f+y;x.f_0,x.f_1,x.f_2;x.e).$$

∎

Proposition 6.5 (subtraction of a double from an AffApprox): If x is an AffApprox and y is a double, then $S(x-y)\supseteq S(x)-S(y)$, where

$$x-y\equiv(x.f-y;x.f_0,x.f_1,x.f_2;x.e).$$

∎

Proposition 6.6 (multiplication): If x and y are AffApprox's, then $S(x \times y) \supseteq S(x) \times S(y)$, where

$$x \times y \equiv (x.f \times y.f; x.f \times y.f_0 + x.f_0 \times y.f, x.f \times y.f_1 + x.f_1 \times y.f, x.f \times y.f_2 + x.f_2 \times y.f; \\ (size(x) + x.e) \times (size(y) + y.e) + (|x.f| \times y.e + x.e \times |y.f|))$$

with $size(x) = |x.f_0| + |x.f_1| + |x.f_2|$ and $size(y) = |y.f_0| + |y.f_1| + |y.f_2|$

Proof: If $g \in S(x)$ and $h \in S(y)$ then we must show that $g \times h \in S(x \times y)$. That is, we need to show

$$|(g \times h)(z_0, z_1, z_2) - ((x.f \times y.f)+ \\ (x.f \times y.f_0 + x.f_0 \times y.f)z_0 + (x.f \times y.f_1 + x.f_1 \times y.f)z_1 + (x.f \times y.f_2 + x.f_2 \times y.f)z_2)| \\ \leq (size(x) + x.e) \times (size(y) + y.e) + (|x.f| \times y.e + x.e \times |y.f|)$$

Note that for any point $(z_0, z_1, z_2) \in A$ and any functions $g \in S(x)$ and $h \in S(y)$ we can find complex numbers u, v with $|u| \leq 1$ and $|v| \leq 1$, such that

$$g(z_0, z_1, z_2) = x.f + (x.f_0 z_0 + x.f_1 z_1 + x.f_2 z_2) + (x.e)u$$

and

$$h(z_0, z_1, z_2) = y.f + (y.f_0 z_0 + y.f_1 z_1 + y.f_2 z_2) + (y.e)v.$$

Multiplying out, we see that

$$(g \times h)(z_0, z_1, z_2) = (x.f \times y.f)+ \\ (x.f \times y.f_0 + x.f_0 \times y.f)z_0 + (x.f \times y.f_1 + x.f_1 \times y.f)z_1 + (x.f \times y.f_2 + x.f_2 \times y.f)z_2+ \\ (x.f \times y.e)v + (x.e \times y.f)u+ \\ ((x.f_0 z_0 + x.f_1 z_1 + x.f_2 z_2) + (x.e)u) \times ((y.f_0 z_0 + y.f_1 z_1 + y.f_2 z_2) + (y.e)v)$$

Hence,

$$|(g \times h)(z_0, z_1, z_2) - ((x.f \times y.f)+ \\ ((x.f \times y.f_0 + x.f_0 \times y.f)z_0 + (x.f \times y.f_1 + x.f_1 \times y.f)z_1 + (x.f \times y.f_2 + x.f_2 \times y.f)z_2))|$$

$$\leq (|x.f|y.e + x.e|y.f|) + (size(x) + x.e) \times (size(y) + y.e).$$

∎

Proposition 6.7 (an AffApprox multiplied by a double): If x is an AffApprox and y is a double, then $S(x \times y) \supseteq S(x) \times S(y)$, where

$$x \times y \equiv (x.f \times y; x.f_0 \times y, x.f_1 \times y, x.f_2 \times y; x.e \times |y|)$$

$\blacksquare$

Proposition 6.8 (division): If x and y are AffApprox's with $|y.f| > size(y) + y.e$, then $S(x/y) \supseteq S(x)/S(y)$, where

$$x/y \equiv (x.f/y.f; (-x.f \times y.f_0 + x.f_0 \times y.f)/((y.f)^2), (-x.f \times y.f_1 + x.f_1 \times y.f)/((y.f)^2),$$
$$(-x.f \times y.f_2 + x.f_2 \times y.f)/((y.f)^2);$$
$$(|x.f| + size(x) + x.e)/(|y.f| - (size(y) + y.e)) -$$
$$((|x.f|/|y.f| + size(x)/|y.f|) + |x.f|size(y)/(|y.f||y.f|)))$$

Proof: For notational convenience, denote $(x.f_0 z_0 + x.f_1 z_1 + x.f_2 z_2)$ by $x.f_k z_k$ and similarly for $y.f_k z_k$ and so on. As above, note that for any point $(z_0, z_1, z_2) \in A$ and any functions $g \in S(x)$ and $h \in S(y)$ we can find complex numbers u, v with $|u| \leq 1$ and $|v| \leq 1$, such that

$$g(z_0, z_1, z_2) = x.f + (x.f_k z_k) + (x.e)u$$

and

$$h(z_0, z_1, z_2) = y.f + (y.f_k z_k) + (y.e)v.$$

We compare $(g/h)(z_0, z_1, z_2)$ with its putative affine approximation. That is, we analyze

$$|(x.f + (x.f_k z_k) + (x.e)u)/(y.f + (y.f_k z_k) + (y.e)v) -$$
$$((x.f/y.f) + \frac{(x.f_k)y.f - x.f(y.f_k)}{(y.f)^2} z_k)|$$

Putting this over a common denominator of $|((y.f)^2)(y.f + (y.f_k z_k) + (y.e)v)|$ and cancelling equal terms (in the numerator) we are left with a quotient whose numerator is

$$|x.e((y.f)^2)u - (x.f_k)y.f(y.f_k)z_k - x.f((y.f_k)^2)z_k +$$
$$(x.f)y.f(y.e)v + x.f_k(y.f)y.e(v)z_k - x.f(y.f_k)y.e(v)z_k|.$$

We must show this (first) quotient is bounded by

$$(|x.f| + size(x) + x.e)/(|y.f| - (size(y) + y.e)) -$$
$$((|x.f|/|y.f| + size(x)/|y.f|) + |x.f|size(y)/(|y.f||y.f|)).$$

Putting this over a common denominator of $|(y.f)|^2(|y.f| - (size(y) + y.e))$ and cancelling equal terms (in the numerator) we are left with a second quotient, whose numerator is

$$x.e|y.f|^2 - (-|x.f||y.f|y.e - size(x)|y.f|(size(y) + y.e) - |x.f|size(y)(size(y) + y.e))$$

and we see that all terms in this numerator are positive. Further, the terms in the numerators of the first and second quotients correspond in a natural way, and each term in the numerator of the second quotient is greater than or equal to the absolute value of its corresponding term in the numerator of the first quotient.

Finally, because the denominator in the second quotient is less than or equal to the absolute value of the denominator in the first quotient, we see that the absolute value of the first quotient is less than or equal to the second quotient, as desired. ∎

Proposition 6.9 (division of a double by an AffApprox): If x is a double and y is an AffApprox with $|y.f| > size(y) + y.e$, then $S(x/y) \supseteq S(x)/S(y)$, where

$$x/y \equiv (x/y.f; -x \times y.f_0/((y.f)^2), -x.f \times y.f_1/((y.f)^2), -x.f \times y.f_2/((y.f)^2); \\ (|x|/(|y.f| - (size(y) + y.e)) - \\ (|x|/|y.f| + |x|size(y)/(|y.f||y.f|)))$$

∎

Proposition 6.10 (division of an AffApprox by a double): If x is an AffApprox and y is a double with $|y| > 0$, then $S(x/y) \supseteq S(x)/S(y)$, where

$$x/y \equiv (x.f/y; x.f_0/y, x.f_1/y, x.f_2/y; x.e/|y|)$$

∎

Finally, we do the square root.

Proposition 6.11 (square root): If x is an AffApprox with $|x.f| > size(x) + x.e$, then $S(\sqrt{x}) \supseteq \sqrt{S(x)}$, where

$$\sqrt{x} = (\sqrt{x.f}; \frac{x.f_0}{2\sqrt{x.f}}, \frac{x.f_1}{2\sqrt{x.f}}, \frac{x.f_2}{2\sqrt{x.f}}; \\ \sqrt{|x.f|} - (\frac{size(x)}{2\sqrt{|x.f|}} + \sqrt{|x.f| - (size(x) + x.e)}))$$

If $|x.f| \le size(x) + x.e$ then we use the crude estimate $(0; 0, 0, 0; \sqrt{|x.f| + size(x) + x.e})$.

The branch of the square root of a complex number is determined by the construction of the square root of a complex in Proposition 7.14. In fact, the square root is in the first or fourth quadrant.

Proof: As above, note that for any point $(z_0, z_1, z_2) \in A$ and any function $g \in S(x)$ we can find a complex number u with $|u| \le 1$, such that

$$g(z_0, z_1, z_2) = x.f + (x.f_k z_k) + (x.e)u.$$

Also, because $|x.f| > size(x) + x.e.$ we see that the argument of $x.f + (x.f_k z_k) + (x.e)u$ is within $\pi/2$ of the argument of $x.f$, and therefore, we can require that $\sqrt{g(z_0, z_1, z_2)}$ has argument within $\pi/4$ of the argument of $\sqrt{x.f}$.

We need to show that

$$|\sqrt{x.f + x.f_k z_k + (x.e)u} - (\sqrt{x.f} + \frac{x.f_k z_k}{2\sqrt{x.f}})|$$

$$\leq \sqrt{|x.f|} - (\frac{size(x)}{2\sqrt{|x.f|}} + \sqrt{|x.f| - (size(x) + x.e)}\ )$$

Or, after multiplying both sides by $\sqrt{|x.f|}$,

$$|\sqrt{x.f(x.f + x.f_k z_k + (x.e)u)} - (x.f + (x.f_k)z_k/2)|$$

$$\leq (|x.f| - size(x)/2) - \sqrt{|x.f|(|x.f| - (size(x) + x.e))}$$

The two sides of the inequality are of the form $A - B$ and $C - D$, and we "simplify" by multiplying by $\frac{A+B}{A+B}$ and $\frac{C+D}{C+D}$. We now show that the (absolute value of the) left-hand numerator is less than or equal to the right-hand numerator. Later, we will show that the (absolute value of the) left-hand denominator is larger than or equal to the right-hand denominator. The left-hand numerator is

$$|x.f(x.f + x.f_k z_k + (x.e)u) - (x.f + (x.f_k)z_k/2)^2|$$

$$= |(x.f)^2 + x.f(x.f_k)z_k + x.f(x.e)u - (x.f)^2 - x.f(x.f_k)z_k - ((x.f_k)^2)(z_k)^2/4|$$

$$= |x.f(x.e)u - ((x.f_k)^2)(z_k)^2/4|$$

The right-hand numerator is

$$(|x.f| - size(x)/2)^2 - |x.f|(|x.f| - (size(x) + x.e))$$

$$= |x.f|^2 - |x.f|size(x) + size(x)^2/4 - |x.f|^2 + |x.f|size(x) + |x.f|x.e$$

$$= |x.f|x.e + size(x)^2/4$$

So the left-hand numerator is indeed less than or equal to the right-hand numerator.

We now compare the denominators, but only after dividing each by $\sqrt{|x.f|}$. The left-hand denominator is

$$|\sqrt{x.f + x.f_k z_k + (x.e)u} + (\sqrt{x.f} + \frac{x.f_k z_k}{2\sqrt{x.f}})|$$

while the right-hand denominator is

$$\sqrt{|x.f|} - \frac{size(x)}{2\sqrt{|x.f|}} + \sqrt{|x.f| - (size(x) + x.e)}$$

The claim that the left-hand denominator is greater than or equal to the right-hand denominator is a bit complicated. First, compare the $\sqrt{x.f}$ term and the $\sqrt{|x.f|}$ terms. They are the same distance from the origin. Next, note that as z_k and u take on all values,

$x.f + x.f_k z_k + (x.e)u$ describes a disk centered at $x.f$ and whose radius is less than $|\sqrt{x.f}|$. Hence, $\sqrt{x.f + x.f_k z_k + (x.e)u}$ describes a convex set containing $\sqrt{x.f}$. This set is symmetric about the line joining the origin and $\sqrt{x.f}$. Further, $\sqrt{x.f}+\sqrt{x.f + x.f_k z_k + (x.e)u}$ describes a convex set containing $2\sqrt{x.f}$. This set is also symmetric about the line joining the origin and $\sqrt{x.f}$. It is easy enough to see that no points on this convex symmetric set get closer to the origin than $\sqrt{|x.f|} + \sqrt{|x.f| - (size(x) + x.e)}$.

Finally, because $|\frac{x.f_k z_k}{2\sqrt{x.f}}| \leq \frac{size(x)}{2\sqrt{|x.f|}}$, no points of

$$\sqrt{x.f} + \sqrt{x.f + x.f_k z_k + (x.e)u} + \frac{x.f_k z_k}{2\sqrt{x.f}}$$

can get closer to the origin than

$$\sqrt{|x.f|} + \sqrt{|x.f| - (size(x) + x.e)} - \frac{size(x)}{2\sqrt{|x.f|}}$$

∎

Example 6.7 (Continuation of Example 5.4): We can now complete the analysis begun in Example 5.4, because we can decribe f and w as 2-by-2 matrices of AffApprox's. For convenience, we repeat the description of the sub-box under investigation.

The sub-box $Z(s01011)$ with

$$s = 00100011000111011100111100010111111110111110011100111100000111101111 0111$$

is the region where

$$\begin{pmatrix} -1.381589027741\ldots \leq Re(L') \leq -1.379848991182\ldots \\ -1.378124546093\ldots \leq Re(D') \leq -1.376574349753\ldots \\ 0.999893182771\ldots \leq Re(R') \leq 1.001274250703\ldots \\ -2.535837191243\ldots \leq Im(L') \leq -2.534606799593\ldots \\ 2.535404997792\ldots \leq Im(D') \leq -2.534308843448\ldots \\ -0.001953125000\ldots \leq Im(R') \leq 0.000000000000\ldots \end{pmatrix}$$

For this sub-box, we get (printing only 10 decimal places, for visual convenience)

$$f = \begin{bmatrix} \begin{pmatrix} -0.8677851121 + i1.4607429651; \\ 0.0000248810 - i0.0003125810, \\ 0.0000000000 + i0.0000000000, \\ 0.0000000000 + i0.0000000000; \\ 0.0000000289 \end{pmatrix} & \begin{pmatrix} 0.0000000000 + i0.0000000000; \\ 0.0000000000 + i0.0000000000, \\ 0.0000000000 + i0.0000000000, \\ 0.0000000000 + i0.0000000000; \\ 0.0000000000 \end{pmatrix} \\ \begin{pmatrix} 0.0000000000 + i0.0000000000; \\ 0.0000000000 + i0.0000000000, \\ 0.0000000000 + i0.0000000000, \\ 0.0000000000 + i0.0000000000; \\ 0.0000000000 \end{pmatrix} & \begin{pmatrix} -0.3006023265 - i0.5060039953; \\ -0.0000909686 - i0.0000593570, \\ 0.0000000000 + i0.0000000000, \\ 0.0000000000 + i0.0000000000; \\ 0.0000000301 \end{pmatrix} \end{bmatrix}$$

and

$$w = \begin{bmatrix} \begin{pmatrix} -0.5845111829 + i0.4773282853; \\ 0.0000000000 + i0.0000000000, \\ -0.0000296707 - i0.0001657332, \\ -0.0004345111 - i0.0001209539; \\ 0.0000002590 \end{pmatrix} & \begin{pmatrix} -0.2840228472 + i0.9825063583; \\ 0.0000000000 + i0.0000000000, \\ 0.0000516606 - i0.0001128245, \\ 0.0005776611 - i0.0001998632; \\ 0.0000006462 \end{pmatrix} \\ \begin{pmatrix} -0.2832291572 + i0.9833572297; \\ 0.0000000000 + i0.0000000000, \\ 0.0000515806 - i0.0001129408, \\ -0.0005778031 + i0.0002005440; \\ 0.0000002806 \end{pmatrix} & \begin{pmatrix} -0.5846352333 + i0.4764792236; \\ 0.0000000000 + i0.0000000000, \\ -0.0000294917 - i0.0001656653, \\ 0.0004341392 + i0.0001213070; \\ 0.0000005286 \end{pmatrix} \end{bmatrix}.$$

Calculating $g = f^{-1}wf^{-1}w^{-1}f^{-1}w^{-1}fw^{-1}f^{-1}w^{-1}f^{-1}wf^{-1}wfww$ gives

$$g = \begin{bmatrix} \begin{pmatrix} -0.5764337542 + i0.4752708071; \\ -0.0031657223 - i0.0001436786, \\ -0.0017723577 + i0.0000352928, \\ -0.0011623491 + i0.0017516088; \\ 0.0008229225 \end{pmatrix} & \begin{pmatrix} -0.2704033973 + i0.9822741250; \\ -0.0045902952 - i0.0019135041, \\ -0.0026219461 - i0.0007506230, \\ -0.0002823450 + i0.0033805602; \\ 0.0008037640 \end{pmatrix} \\ \begin{pmatrix} -0.2861207992 + i0.9766064999; \\ -0.0002777968 + i0.0020330488, \\ 0.0000837571 + i0.0010241875, \\ 0.0028322367 - i0.0005972336; \\ 0.0018172437 \end{pmatrix} & \begin{pmatrix} -0.5861133046 + i0.4624368851; \\ -0.0021932627 + i0.0040523411, \\ -0.0008612361 + i0.0022394639, \\ 0.0061581377 - i0.0005862070; \\ 0.0017738513 \end{pmatrix} \end{bmatrix}.$$

We then get

$$length(g) = \begin{pmatrix} -1.3588762105 - i2.4897230182; \\ 0.0030210500 - i0.0182284729, \\ 0.0007938572 - i0.0096614614, \\ -0.0122034521 + i0.0074353043; \\ 0.0080071969 \end{pmatrix}$$

and

$$\frac{length(g)}{L'} = \begin{pmatrix} 0.9825397896 - i0.0008933519; \\ 0.0053701602 + i0.0037789019, \\ 0.0028076072 + i0.0018421952, \\ -0.0002400615 - i0.0049443045; \\ 0.0027802966 \end{pmatrix}.$$

This isn't quite good enough to kill the sub-box, since $|length(g)/L'|$ can be high as 1.0001951323.

When we subdivide $Z(s01011)$, we have to analyze two sub-boxes, $Z(s010110)$ and $Z(s010111)$. For $Z(s010110)$, the same calculation on the region

$-1.381589027741073400 \leq Re(L') \leq -1.379848991182205200$

$-1.378124546093485700 \leq Re(D') \leq -1.376574349753672900$

$0.999893182771602220 \leq Re(R') \leq 1.001274250703607400$

$-2.535837191243490300 \leq Im(L') \leq -2.534606799593201600$
$-2.535404997792558600 \leq Im(D') \leq -2.534308843448505900$
$-0.001953125000000000 \leq Im(R') \leq -0.000976562500000000$
gives

$$\frac{length(g)}{L'} = \begin{pmatrix} 0.9814518667 + i0.0008103446; \\ 0.0053616729 + i0.0037834001, \\ 0.0028027236 + i0.0018435245, \\ -0.0013175066 - i0.0032448794; \\ 0.0019033926 \end{pmatrix},$$

and we can then bound $|\frac{length(g)}{L'}| \leq 0.9967745579$, which kills $Z(s010110)$.

On $Z(s010111)$, the calculation gives

$$\frac{length(g)}{L'} = \begin{pmatrix} 0.9836225919 - i0.0025990177; \\ 0.0053786346 + i0.0037743930, \\ 0.0028124892 + i0.0018408583, \\ -0.0013333182 - i0.0032343347; \\ 0.0019044429 \end{pmatrix}$$

and $|\frac{length(g)}{L'}| \leq 0.9989610507$, which kills $Z(s010111)$.

Chapter 7: Complex Numbers with Round-Off Error

Remark 7.1: The theoretical method for proving Theorem 0.2 has been implemented via the computer programs *verify* and *corona*, which are available, together with the relevant data sets, at the *Annals* web site. To make this computer-aided proof rigorous, we needed to deal with round-off error in calculations.

One approach to round-off error would be to use interval arithmetic packages to carry out all calculations with floating-point numbers (also called "doubles"), or to generate our own version of these packages. However, it appears that this would be much too slow given the size of our collection of sub-boxes and conditions/killerwords.

To solve this problem of speed, we implement round-off error at a higher level of programing. That is, we incorporate round-off error directly into AffApprox's, which makes our error calculations more accurate, thereby avoiding much sub-division of sub-boxes. This necessitates that we incorporate round-off error directly into complex numbers as well.

Definition 7.2: In the next Chapter we work with AffApprox's. In this section we show how to do standard operations on complex numbers while keeping track of round-off error. There are two types of complex numbers to consider:

1.) An *XComplex* $x = (x.re, x.im)$ corresponds to a complex number that is represented exactly; it simply consists of a real part and an imaginary part.

2.) An *AComplex* $x = (x.re, x.im; x.e)$ corresponds to an "interval" that contains the complex number in question. Thus, it consists of an XComplex and a floating-point number representing the error. In particular, the AComplex x represents the set $S(x)$ of complex numbers $\{w \ : \ |w - (x.re + i(x.im))| \leq x.e \ \}$.

Remark 7.3: In general, our operations act on XComplexes and produce AComplexes, or they act on AComplexes and produce AComplexes . In one case, the unary minus, an XComplex goes to an XComplex. In the calculations that follow the effect on the error is the whole point.

Conventions and Standards 7.4: We begin, by writing down our basic rules, which follow easily from the IEEE-754 standard for machine arithmetic (see [IEEE]). (Actually, the "hypot" function $h(a, b)$, which computes by elaborate chicanery $\sqrt{a^2 + b^2}$, is not part of the IEEE-754 standard, but satisfies the appropriate standard according to the documentation provided (see [K1]).) The operations here are on double-precision floating-point real numbers ("doubles") and we denote a true operation by the usual symbol and the associated machine operation by the same symbol in a circle, with two exceptions: a machine square root $\sqrt{a}$ is denoted $\sqrt[\circ]{a}$ and the machine version of the hypot function is denoted $h_\circ$. Perhaps a third exception is our occasional notation of true multiplication by the absence of a symbol.

There is a finite set of numbers (sometimes called "machine numbers") which are representable on the computer. Ignoring technicalities, a non-zero floating-point number is represented by a fixed number of bits of which the first determines the sign of the number, the next m represent the exponent, and the remaining n represent the mantissa of the number. Because our non-zero numbers start with a 1, that means the n mantissa bits actually represent the next n binary digits after the 1. That is, the mantissa is actually $1.b_1b_2b_3...b_n$. The IEEE-754 standard calls for 64-bit doubles with $m = 11$ and $n = 52$. We define EPS to be 2^{-n}, in which case $EPS/2$ is $2^{-(n+1)}$.

The IEEE-754 standard states that the result of an operation is always the closest representable number to the true solution (as long as we are in the bounds of representable numbers). For example, for machine numbers a and b, we have $a \oplus b = m(a + b)$ where m is the function which take the machine value of its argument (when it lies in the range of representable numbers). Thus, properties of the type

$$|(a + b) - (a \oplus b)| \le (EPS/2)|a + b|$$

follow immediately from the IEEE-754 standard, as long as we do not *underflow* or *overflow* outside of the range of representable numbers.

Specifically, underflow occurs when the result of an operation is smaller in absolute value than 2^{-1023}, and overflow occurs when the result of an operation is larger in absolute value than roughly 2^{1025} (see [IEEE] Section 7).

We further note that the formula

$$|(a + b) - (a \oplus b)| \le (EPS/2)|a \oplus b|$$

follows because the true answer has "exponent" which is less than or equal to the exponent of the machine answer. We reiterate, that in both cases, a and b are assumed to be machine numbers.

Of course, a machine operation such as $\oplus$ must act on doubles, while a "true" operation such as $+$ can act on reals (which includes doubles). In this chapter, long strings of inequalities will be used to prove the various propositions, and care was taken to ensure that machine operations act on machine numbers. In particular, the various variables appearing in the propositions are assumed to be doubles. The IEEE-754 standard provides for conversions from decimal to binary (within the appropriate range, conversion is to the nearest representable number) and from binary to decimal. However, these are rarely used in this paper, although a trivial class of exceptions is provided by the decimal numbers in the conditions of Chapter 5.

When calculations underflow or overflow outside of the range of representable numbers, we require that the computer inform us if either exception has occurred.

As in Chapter 6, we now break with the usual numbering convention. Note that the above comments provide a proof of the following properties.

Basic Properties 7.0 (assuming no underflow and no overflow):

In the formulas that follow, a,b, and A are machine numbers.

$1+k\times EPS = 1\oplus(k\otimes EPS)$ when k is an integer which is not huge in absolute value (that is, smaller than roughly 2^{50}). Thus, within the appropriate range, $1+k\times EPS$ is a machine number.

$2^k\times A = 2^k\otimes A$ when k is an integer and $2^k\otimes A$ neither underflows nor overflows.

$$|(a+b)-(a\oplus b)| \le (EPS/2)|a+b|$$

$$|(a+b)-(a\oplus b)| \le (EPS/2)|a\oplus b|$$

Analagous formulas hold for $-,\ *,\ /,\ \surd.$

$$|h(a,b)-h_\circ(a,b)| \le (EPS)|h(a,b)|$$

$$|h(a,b)-h_\circ(a,b)| \le (EPS)|h_\circ(a,b)|$$

¿From these formulas, we immediately compute the following.

$$(1-EPS/2)|a+b| \le |a\oplus b| \le (1+EPS/2)|a+b|$$

$$(1-EPS/2)|a\oplus b| \le |a+b| \le (1+EPS/2)|a\oplus b|$$

Analagous formulas hold for $-,\ *,\ /,\ \surd.$

$$(1-EPS)|h(a,b)| \le |h_\circ(a,b)| \le (1+EPS)|h(a,b)|$$

$$(1-EPS)|h_\circ(a,b)| \le |h(a,b)| \le (1+EPS)|h_\circ(a,b)|$$

Of course, we can also get the following type of formula, which is sometimes convenient, for example, in the proof of Lemma 7.2.

$$(\frac{1}{1+\frac{EPS}{2}})|a \oplus b| \leq |a+b| \leq (\frac{1}{1-\frac{EPS}{2}})|a \oplus b|$$

Before starting in with our Propositions, we prove a couple of lemmas.

Lemma 7.0 0 (assuming no underflow and no overflow): For machine numbers a and b,

$$(1 - EPS) \otimes |a \oplus b| \leq |a+b| \leq (1 + EPS) \otimes |a \oplus b|$$

Analagous formulas hold for $-,\ *,\ /,\ \surd.$

Proof: Assume $a + b > 0$. If $(1 + EPS) \otimes (a \oplus b) < (a + b)$ then the machine number $(1 + EPS) \otimes (a \oplus b)$ is a better approximation to $a + b$ than $a \oplus b$, because $(a \oplus b) < (1 + EPS) \otimes (a \oplus b)$. This contradicts the IEEE standard. The case $a + b < 0$ can be handled similarly, and the case $a + b = 0$ is trivial. Similarly for the left-hand inequality. ∎

Lemma 7.1:

$$(1 + EPS/2)^a A \leq (1 + kEPS) \otimes A$$

where A *is a non-negative machine number, and* a *is a (not huge) integer, such that for* a *even,* $k = \frac{a}{2} + 1$ *and for* a *odd,* $k = \frac{a+1}{2} + 1$.

Proof:

$$(1 + EPS/2)^a A \leq (1 - EPS/2)(1 + kEPS)A \leq (1 + kEPS) \otimes A$$

The first inequality holds if a and k are as in the Lemma, and the second inequality is a consequence of one of the formulas preceding Lemma 7.0 ($A \geq 0$). ∎

We now begin our construction of complex arithmetic. We will give proofs for most of the operations; the others should be straightforward to derive, or can be found in the *Annals* archive.

Remarks 7.5: i) We remind the reader that all machine operations are on machine numbers, and that the various variables appearing in the propositions are assumed to be doubles.

ii) The propositions that follow include in their statements the definitions of the various operations (see Remark 6.6iii).

Proposition 7.1 (-X): If x is an XComplex, then $S(-x) = -S(x)$, where

$$-x \equiv (-x.re, -x.im)$$

∎

Proposition 7.2 (X + D): If x is an XComplex and d is a double, then $S(x+d) \supseteq S(x) + S(d)$, where

$$x + d \equiv (x.re \oplus d, x.im; (EPS/2) \otimes |x.re \oplus d|)$$

Proof: The error is bounded by

$$|(x.re + d) - (x.re \oplus d)| \leq (EPS/2)|x.re \oplus d| = (EPS/2) \otimes |x.re \oplus d|$$

∎

Proposition 7.3 (X - D): If x is an XComplex and d is a double, then $S(x-d) \supseteq S(x) - S(d)$, where

$$x - d \equiv (x.re \ominus d, x.im; (EPS/2) \otimes |x.re \ominus d|)$$

∎

Proposition 7.4 (X + X): If x and y are XComplex's, then $S(x+y) \supseteq S(x) + S(y)$, where

$$x + y \equiv (x.re \oplus y.re, x.im \oplus y.im; (EPS/2) \otimes ((1+EPS) \otimes (|x.re \oplus y.re| \oplus |x.im \oplus y.im|)))$$

Proof: The error is bounded by

$$\begin{aligned} &|(x.re + y.re) - (x.re \oplus y.re)| + |(x.im + y.im) - (x.im \oplus y.im)| \\ &\leq (EPS/2)(|x.re \oplus y.re| + |x.im \oplus y.im|) \\ &\leq (EPS/2)((1+EPS) \otimes (|x.re \oplus y.re| \oplus |x.im \oplus y.im|)) \\ &= (EPS/2) \otimes ((1+EPS) \otimes (|x.re \oplus y.re| \oplus |x.im \oplus y.im|)) \end{aligned}$$

To go from line 2 to line 3 we used Lemma 7.0. ∎

Proposition 7.5 (X - X): If x and y are XComplex's, then $S(x-y) \supseteq S(x) - S(y)$, where

$$x - y \equiv (x.re \ominus y.re, x.im \ominus y.im; (EPS/2) \otimes ((1+EPS) \otimes (|x.re \ominus y.re| \oplus |x.im \ominus y.im|)))$$

∎

Proposition 7.6 (A + A): If x and y are AComplex's, then $S(x+y) \supseteq S(x) + S(y)$, where

$x + y \equiv (re, im; e)$ with
$re = x.re \oplus y.re$

$im = x.im \oplus y.im$
$e = (1 + 2EPS) \otimes (((EPS/2) \otimes (|re| \oplus |im|)) \oplus (x.e \oplus y.e))$

Proof: The error is bounded by the sum of the contributions from the real part, the imaginary part, and the two individual errors:

$$|(x.re \oplus y.re) - (x.re + y.re)| + |(x.im \oplus y.im) - (x.im + y.im)| + (x.e + y.e).$$

$$\leq (EPS/2)|x.re \oplus y.re| + (EPS/2)|x.im \oplus y.im| + (1 + EPS/2)(x.e \oplus y.e)$$

$$\leq (1 + EPS/2)(EPS/2)(|x.re \oplus y.re| \oplus |x.im \oplus y.im|) + (1 + EPS/2)(x.e \oplus y.e)$$

$$= (1 + EPS/2)((EPS/2)(|x.re \oplus y.re| \oplus |x.im \oplus y.im|) + (x.e \oplus y.e))$$

$$\leq (1 + EPS/2)^2(((EPS/2)(|x.re \oplus y.re| \oplus |x.im \oplus y.im|)) \oplus (x.e \oplus y.e))$$

$$\leq (1 + 2EPS) \otimes (((EPS/2) \otimes (|x.re \oplus y.re| \oplus |x.im \oplus y.im|)) \oplus (x.e \oplus y.e))$$

∎

The precedence for machine operations is the same as that for true operations, so one pair of parentheses is unnecessary and will often be omitted in what follows.

Proposition 7.7 (A - A): If x and y are AComplex's, then $S(x - y) \supseteq S(x) - S(y)$, where
$x - y \equiv (re, im; e)$ with
$re = x.re \ominus y.re$
$im = x.im \ominus y.im$
$e = (1 + 2EPS) \otimes (((EPS/2) \otimes (|re| \oplus |im|)) \oplus (x.e \oplus y.e))$ ∎

Proposition 7.8 (X × D): If x is an XComplex and d is a double, then $S(x \times d) \supseteq S(x) \times S(d)$, where
$x \times d \equiv (re, im; e)$ with
$re = x.re \otimes d$
$im = x.im \otimes d$
$e = (EPS/2) \otimes ((1 + EPS) \otimes (|re| \oplus |im|))$ ∎

Proposition 7.9 (X / D): If x is an XComplex and d is a double, then $S(x/d) \supseteq S(x)/S(d)$, where
$x/d \equiv (re, im; e)$ with
$re = x.re \oslash d$
$im = x.im \oslash d$
$e = (EPS/2) \otimes ((1 + EPS) \otimes (|re| \oplus |im|))$ ∎

Proposition 7.10 (X × X): If x and y are XComplex's, then $S(x \times y) \supseteq S(x) \times S(y)$, where
$x \times y \equiv (re, im; e)$ with
$re = re1 \ominus re2$, with $re1 = x.re \otimes y.re$ and $re2 = x.im \otimes y.im$

$im = im1 \oplus im2$, with $im1 = x.re \otimes y.im$ and $im2 = x.im \otimes y.re$
$e = EPS \otimes ((1 + 2EPS) \otimes ((|re1| \oplus |re2|) \oplus (|im1| \oplus |im2|)))$

Proof: The error is bounded by the sum of the contributions from the real part and the imaginary part:

$$|(x.re \times y.re - x.im \times y.im) - ((x.re \otimes y.re) \ominus (x.im \otimes y.im))| + |(x.re \times y.im + x.im \times y.re) - ((x.re \otimes y.re) \oplus (x.im \otimes y.im))|$$

We want to bound this by a machine formula. Let's begin by bounding

$$|(x.re \times y.re - x.im \times y.im) - ((x.re \otimes y.re) \ominus (x.im \otimes y.im))|$$

by a machine formula.

$$|(x.re \times y.re - x.im \times y.im) - ((x.re \otimes y.re) \ominus (x.im \otimes y.im))|$$

$$\leq |((x.re \times y.re) - (x.im \times y.im)) - ((x.re \otimes y.re) - (x.im \otimes y.im))| + |((x.re \otimes y.re) - (x.im \otimes y.im)) - ((x.re \otimes y.re) \ominus (x.im \otimes y.im))|$$

$$\leq |(x.re \times y.re) - (x.re \otimes y.re)| + |(x.im \times y.im) - (x.im \otimes y.im)| + (EPS/2)|(x.re \otimes y.re) - (x.im \otimes y.im)|$$

$$\leq (EPS/2)|(x.re \otimes y.re)| + (EPS/2)|(x.im \otimes y.im)| + (EPS/2)(|x.re \otimes y.re| + |x.im \otimes y.im|)$$

$$= (EPS/2)(2)(|x.re \otimes y.re| + |x.im \otimes y.im|)$$

$$\leq EPS(1 + EPS/2)(|x.re \otimes y.re| \oplus |(x.im \otimes y.im|)$$

Almost the exact same calculation produces the analagous formula for the imaginary contribution, and we now combine the two to get a bound on the total error.

$$\leq EPS(1 + EPS/2)(|x.re \otimes y.re| \oplus |x.im \otimes y.im|) + EPS(1 + EPS/2)(|x.re \otimes y.im| \oplus |x.im \otimes y.re|)$$

$$\leq EPS \otimes ((1 + 2EPS) \otimes ((|x.re \otimes y.re| \oplus |x.im \otimes y.im|) \oplus (|x.re \otimes y.im| \oplus |x.im \otimes y.re|)))$$

∎

Proposition 7.11 (D / X): If x is a double and y is an XComplex, then $S(x/y) \supseteq S(x)/S(y)$, where

$x/y \equiv (re, im; e)$ with
$re = (x \otimes y.re) \oslash nrm$ where $nrm = y.re \otimes y.re \oplus y.im \otimes y.im$
$im = -(x \otimes y.im) \oslash nrm$
$e = (2EPS) \otimes ((1 + 2EPS) \otimes (|re| \oplus |im|))$

Proof: The true version of x/y is equal to $(x \times y.re + i(-x \times y.im))/((y.re)^2 + (y.im)^2)$ and we need to compare this with the machine version to find the error. Further, this error is less than or equal to the sum of the real error and the imaginary error. Thus, we start with the real calculation (as in the statement of the proposition, we use nrm to represent the machine version of $(y.re)^2 + (y.im)^2$).

$$|\frac{x \times y.re}{(y.re)^2 + (y.im)^2} - ((x \otimes y.re) \oslash nrm)|$$

$$\leq |(x \otimes y.re) \oslash nrm - \frac{x \otimes y.re}{nrm}| + |\frac{x \otimes y.re}{nrm} - \frac{x \times y.re}{nrm}| + |\frac{x \times y.re}{nrm} - \frac{x \times y.re}{(y.re)^2 + (y.im)^2}|$$

Before continuing, let's compare $\frac{1}{nrm}$ and $\frac{1}{(y.re)^2+(y.im)^2}$ by developing a formula for comparing $\frac{1}{a^2+b^2}$ and its associated $\frac{1}{nrm}$:

Lemma 7.2:
$$|\frac{1}{nrm} - \frac{1}{a^2 + b^2}| \leq (EPS + (EPS/2)^2)\frac{1}{nrm}$$
where $nrm = a \otimes a \oplus b \otimes b$.

Proof: We compute that

$$(\frac{1}{1 + EPS/2})^2 \times nrm \leq a^2 + b^2 \leq (\frac{1}{1 - EPS/2})^2 \times nrm,$$

hence

$$\frac{1}{nrm}(1 - EPS/2)^2 \leq \frac{1}{a^2 + b^2} \leq \frac{1}{nrm}(1 + EPS/2)^2.$$

It then follows that

$$|\frac{1}{nrm} - \frac{1}{a^2 + b^2}| \leq \frac{1}{nrm}(1 + EPS/2)^2 - \frac{1}{nrm}$$

$$= \frac{1}{nrm}((1 + EPS/2)^2 - 1) = (EPS + (EPS/2)^2)\frac{1}{nrm}$$

∎

Getting back to our main calculation (with $nrm = y.re \otimes y.re \oplus y.im \otimes y.im$),

$$|(x \otimes y.re) \oslash nrm - \frac{x \otimes y.re}{nrm}| + |\frac{x \otimes y.re}{nrm} - \frac{x \times y.re}{nrm}| + |\frac{x \times y.re}{nrm} - \frac{x \times y.re}{(y.re)^2 + (y.im)^2}|$$

$$\le (EPS/2)\frac{|x \otimes y.re|}{nrm} + (EPS/2)\frac{|x \otimes y.re|}{nrm} + (EPS + (EPS/2)^2)\frac{|x \times y.re|}{nrm}$$

$$= (EPS/2)(\frac{1}{nrm})(2|x \otimes y.re| + (2 + EPS/2) \times |x \times y.re|)$$

$$\le (EPS/2)(\frac{1}{nrm})(2|x \otimes y.re| + (2 + EPS/2)(1 + EPS/2) \times |x \otimes y.re|)$$

$$= (EPS/2)(\frac{1}{nrm})(|x \otimes y.re|)(2 + (2 + EPS/2)(1 + EPS/2))$$

$$\le (EPS/2)(4 + 3EPS/2 + (EPS/2)^2)(|x \otimes y.re|)(\frac{1}{nrm})$$

$$\le (EPS/2)(4 + 3EPS/2 + (EPS/2)^2)(1 + EPS/2)(|x \otimes y.re| \oslash nrm)$$

$$\le (2EPS)(1 + 3EPS/8 + (EPS/4)^2)(1 + EPS/2)(|(x \otimes y.re \oslash nrm)|)$$

We also get the analagous formula for the imaginary contribution for the error, so our total error is bounded by

$$(2EPS)(1+3EPS/8+(EPS/4)^2)(1+EPS/2)((|(x\otimes y.re)\oslash nrm|)+(|(x\otimes y.im)\oslash nrm|))$$

$$\le (2EPS)(1+3EPS/8+(EPS/4)^2)(1+EPS/2)^2((|(x\otimes y.re)\oslash nrm|)\oplus(|(x\otimes y.im)\oslash nrm|))$$

$$\le (2EPS)(1 - EPS/2)(1 + 2EPS)((|(x \otimes y.re) \oslash nrm|) \oplus (|(x \otimes y.im) \oslash nrm|))$$

$$\le (2EPS) \otimes ((1 + 2EPS) \otimes ((|(x \otimes y.re) \oslash nrm|) \oplus (|(x \otimes y.im) \oslash nrm|)))$$

Here we used the fact that

$$(1 + 3EPS/8 + (EPS/4)^2)(1 + EPS/2)^2 \le (1 - EPS/2)(1 + 2EPS)$$

∎

This should give the flavor of division proofs. As such, we will skip the proofs of X/X and A/A and simply refer to the *Annals* archive.

Proposition 7.12 (X / X): If x and y are XComplex's, then $S(x/y) \supseteq S(x)/S(y)$, where
$x/y \equiv (re, im; e)$ with
$re = (x.re \otimes y.re \oplus x.im \otimes y.im) \oslash nrm$ where $nrm = y.re \otimes y.re \oplus y.im \otimes y.im$
$im = (x.im \otimes y.re \ominus x.re \otimes y.im) \oslash nrm$
$e = (5EPS/2) \otimes ((1 + 3EPS) \otimes A)$ where

$$A = ((|x.re \otimes y.re| \oplus |x.im \otimes y.im|) \oplus (|x.im \otimes y.re| \oplus |x.re \otimes y.im|)) \oslash nrm$$

∎

Proposition 7.13 (A / A): If x and y are AComplex's with $y.e < 100EPS \otimes |y|$, or, more accurately,

$$(y.e)^2 < ((10000EPS) \otimes EPS) \otimes nrm$$

then $S(x/y) \supseteq S(x)/S(y)$, where

$x/y \equiv (re, im; e)$ with

$re = (x.re \otimes y.re \oplus x.im \otimes y.im) \oslash nrm$ where $nrm = y.re \otimes y.re \oplus y.im \otimes y.im$

$im = (x.im \otimes y.re \ominus x.re \otimes y.im) \oslash nrm$

$e = (1 + 4EPS) \otimes (((5EPS/2) \otimes A \oplus (1 + 103EPS) \otimes B) \oslash nrm)$ where

$$A = (|x.re \otimes y.re| \oplus |x.im \otimes y.im|) \oplus (|x.im \otimes y.re| \oplus |x.re \otimes y.im|)$$

$$B = x.e \otimes (|y.re| \oplus |y.im|) \oplus (|x.re| \oplus |x.im|) \otimes y.e$$

∎

Our last proposition will construct the square-root function. As a warm-up, ignoring round-off error, our construction is as follows. If $x = x.re + ix.im$ then $\sqrt{x} = s + id$ where $s = \sqrt{(|x.re| + h(x.re, x.im))/2}$ and $d = x.im/(2s)$ when $x.re > 0.0$, and $\sqrt{x} = d + is$ otherwise. Thus, we take our (no-round-off) square roots to be in the first and fourth quadrants.

Proposition 7.14 ($\sqrt{X}$): If x is an XComplex, then $S(\sqrt{x}) \supseteq \sqrt{S(x)}$ where we let $s_o = \sqrt[o]{(|x.re| \oplus h_o(x.re, x.im)) \otimes 0.5}$ and $d_o = (x.im \oslash s) \otimes 0.5$, and define

$\sqrt{x} \equiv (re, im; e)$ where

$re = s_o$ if $x.re > 0.0$ and $re = d_o$ otherwise,

$im = d_o$ if $x.re > 0.0$ and $im = s_o$ otherwise,

$e = EPS \otimes ((1 + 4EPS) \otimes (1.25 \otimes s_o \oplus 1.75 \otimes |d_o|))$

Proof: This will be a little nasty. Let's begin by analyzing e_s, which is the difference between the true calculation of s and the machine calculation of s, that is $e_s = |s - s_o|$. First, we bound s.

$$\begin{aligned} s &= \sqrt{(|x.re| + h(x.re, x.im)) * 0.5} \\ &\leq (1 + EPS)^{1/2} \sqrt{(|x.re| + h_o(x.re, x.im)) * 0.5} \\ &\leq (1 + EPS)^{1/2}(1 + EPS/2)^{1/2} \sqrt{(|x.re| \oplus h_o(x.re, x.im)) * 0.5} \\ &\leq (1 + EPS)^{1/2}(1 + EPS/2)^{1/2}(1 + EPS/2) \sqrt[o]{(|x.re| \oplus h_o(x.re, x.im)) * 0.5} \\ &= (1 + EPS)^{1/2}(1 + EPS/2)^{3/2} s_o \end{aligned}$$

By a power series expansion, we see that

$$(1 + EPS)^{1/2}(1 + EPS/2)^{3/2}$$

$$= (1 + \frac{1}{2}EPS - \frac{1}{8}EPS^2 + ...) + (1 + \frac{3}{2}EPS/2 + \frac{3}{8}(EPS/2)^2 + ...)$$

$$= (1 + \frac{5}{4}EPS + \frac{11}{32}EPS^2 + ...)$$

So that,

$$s \leq (1 + \frac{5}{4}EPS + \frac{11}{32}EPS^2 + ...)s_o$$

Similarly,

$$s \geq (1 - \frac{5}{4}EPS)s_o$$

Thus, we can bound the s error,

$$e_s = |s - s_o|$$

$$\leq ((1 + \frac{5}{4}EPS + \frac{11}{32}EPS^2 + ...) - 1)s_o$$

$$= (\frac{5}{4}EPS + \frac{11}{32}EPS^2 + ...)s_o$$

Next, we analyze e_d, which is the absolute value of the difference between the true calculation of d and the machine calculation of d. That is, $e_d = |d - d_o|$.

$$e_d = |x.im/(2s) - x.im \oslash (2s_o)|$$

$$\leq |x.im \oslash (2s_o) - x.im/(2s_o)| + |x.im/(2s_o) - x.im/(2s)|$$

$$\leq (EPS/2)|x.im/(2s_o)| + |\frac{x.im}{2}\frac{s - s_o}{ss_o}|$$

$$\leq (EPS/2)|x.im/(2s_o)| + |\frac{x.im}{2}\frac{1}{ss_o}((5/4)EPS + (11/32)EPS^2 + ...)s_o|$$

$$\leq (EPS/2)|x.im/(2s_o)| + |\frac{x.im}{2}\frac{1}{s_o(1 - (5/4)EPS)}((5/4)EPS + (11/32)EPS^2 + ...)|$$

$$= (EPS/2)|x.im/(2s_o)|(1 + \frac{(5/2) + (11/16)EPS + ...)}{(1 - (5/4)EPS)})$$

$$= (EPS/2)\frac{(7/2) + (-9/16)EPS + ...}{(1 - (5/4)EPS)}|x.im/(2s_o)|$$

$$\leq (EPS/2)(1 + EPS/2)\frac{7/2}{(1 - (5/4)EPS)}|x.im \oslash (2s_o)|$$

$$= (EPS/2)(1 + EPS/2)\frac{7/2}{(1 - (5/4)EPS)}|d_o|$$

Finally, we can bound the overall error $e = e_s + e_d$.

$$e_s + e_d$$

$$\le (\frac{5}{4}EPS + \frac{11}{32}EPS^2 + ...)s_o + (EPS/2)(1+EPS/2)\frac{7/2}{(1-(5/4)EPS)}|d_o|$$

$$\le (EPS + \frac{11}{40}EPS^2 + ...)(\frac{5}{4}s_o) + EPS(1+EPS/2)\frac{1}{(1-(5/4)EPS)}|\frac{7}{4}d_o|$$

$$\le EPS(1+EPS/2)\frac{1}{(1-(5/4)EPS)}(\frac{5}{4}s_o) + EPS(1+EPS/2)\frac{1}{(1-(5/4)EPS)}|\frac{7}{4}d_o|$$

$$\le EPS(1+EPS/2)\frac{1}{(1-(5/4)EPS)}(\frac{5}{4}s_o + |\frac{7}{4}d_o|)$$

$$\le EPS(1+EPS/2)^3\frac{1}{(1-(5/4)EPS)}(\frac{5}{4} \otimes s_o \oplus |\frac{7}{4} \otimes d_o|)$$

$$\le EPS(1-(EPS/2))(1+4EPS)(\frac{5}{4} \otimes s_o \oplus |\frac{7}{4} \otimes d_o|)$$

$$\le EPS \otimes ((1+4EPS) \otimes (\frac{5}{4} \otimes s_o \oplus |\frac{7}{4} \otimes d_o|))$$

■

Now, we develop a couple of formulas for the absolute value of an XComplex.

Formula 7.0 (absUB(X)):
If x is an XComplex, then we get an upper bound on the absolute value of x as follows.

$$|x| = h(x.re, x.im) \le (1+EPS)h_\circ(x.re, x.im)$$

$$\le (1-EPS/2)(1+2EPS)h_\circ(x.re, x.im)$$

$$\le (1+2EPS) \otimes h_\circ(x.re, x.im)$$

Thus, we define

$$absUB(x) = (1+2EPS) \otimes h_\circ(x.re, x.im)$$

■

Formula 7.1 (absLB(X)):
If x is an XComplex, then we get a lower bound on the absolute value of x as follows.

$$|x| = h(x.re, x.im) \ge (1-EPS)h_\circ(x.re, x.im)$$

$$\ge (1+EPS/2)(1-2EPS)h_\circ(x.re, x.im)$$

$$\ge (1-2EPS) \otimes h_\circ(x.re, x.im)$$

Thus, we define

$$absLB(x) = (1-2EPS) \otimes h_\circ(x.re, x.im)$$

■

Finally, in several places in the programs *verify* and *corona* we perform a standard operation on a pair of doubles and must take into account round-off error. This is easy if we use Lemma 7.0.

For example, in *inequalityHolds* we want to show that $wh \times wh > absUB(along)$, where $wh = absLB(whirle)$. By Lemma 7.0, we know that $(1 - EPS) \otimes (wh \otimes wh) \leq wh \times wh$ and we simply test that $(1 - EPS) \otimes (wh \otimes wh) \geq absUB(along)$.

Similar situations occur in the functions *horizon* and *larger-angle* in *corona*.

A slightly more complicated version of this occurs in the computer calculation of $pos[i]$ and $size[i]$, that is, the center and size of a sub-box. Prior to multiplication by $scale[i] = 2^{(5-i)/6}$, the calculations of *pos* and *size* are exact. However, multiplication by *scale* introduces round-off error. For the center of the box we will have the computer use $pos[i] \otimes scale[i]$ with the realization that this is not necessarily $pos[i] \times scale[i]$. Thus, we have to choose appropriate sizes to ensure that the machine sub-box contains the true sub-box.

Notationally, this is annoying, because we typically use a computer command like $pos[i] = pos[i] \otimes scale[i]$, while in an exposition, we need to avoid that. We will denote the true center of the box by $p[i]$ and the machine center of the box by $p_0[i]$, and the true and machine sizes will be denoted $s[i]$ and $s_0[i]$. We will let $pos[i]$ and $size[i]$ be the position and size (true and machine are the same) before multiplication by $scale[i]$.

Let $p[i] = pos[i] \times scale[i]$, $p_0[i] = pos[i] \otimes scale[i]$, and $s[i] = size[i] \times scale[i]$. We must select $s_0[i]$ so that $p_0[i] + s_0[i] \geq p[i] + s[i]$. (Here, taking $+$ on the left-hand side is correct, because the need for machine calculation there is incorporated at other points in the programs.) So, we must find $s_0[i]$ such that $s_0[i] \geq (p[i] - p_0[i]) + s[i]$.

$$(p[i] - p_0[i]) + s[i]. \leq (EPS/2)|p_0[i]| + size[i] \times scale[i]$$

$$\leq (EPS/2)|p_0[i]| + (1 + EPS/2)(size[i] \otimes scale[i])$$

$$\leq (1 + EPS/2)((EPS/2)|p_0[i]| + (size[i] \otimes scale[i]))$$

$$\leq (1 + EPS/2)^2((EPS/2)|p_0[i]| \oplus (size[i] \otimes scale[i]))$$

$$\leq (1 + 2EPS) \otimes ((EPS/2)|p_0[i]| \oplus (size[i] \otimes scale[i]))$$

Thus we take

$$s_0[i] = (1 + 2EPS) \otimes ((EPS/2)|p_0[i]| \oplus (size[i] \otimes scale[i]))$$

This also works to give $p_0[i] - s_0[i] \leq p[i] - s[i]$. ∎

Chapter 8: AffApprox's with Round-Off Error

In Chapter 6, we saw how to do calculations with AffApprox's. In this chapter, we incorporate round-off error into these calculations.

Conventions 8.1: An AffApprox x is a five-tuple $(x.f; x.f_0, x.f_1, x.f_2; x.err)$ consisting of four complex numbers $(x.f, x.f_0, x.f_1, x.f_2)$ and one real number $x.err$. In Chapter 6, the real number was denoted $x.e$, but it seems preferable to use $x.err$ in this Chapter. Recall (Definition 6.4) that an AffApprox x represents the set $S(x)$ of functions from

$A = \{(z_0, z_1, z_2) \in \mathbf{C}^3 : |z_k| \leq 1 \text{ for } k \in \{0, 1, 2\}\}$ to $\mathbf{C}$ that are $x.err$-well-approximated by the affine function $x.f + x.f_0 z_0 + x.f_1 z_1 + x.f_2 z_2$.

Remark 8.2: A review of Definition 7.2 (XComplex's and AComplex's; loosely, exact and approximate complex numbers) might be helpful at this time.

One approach to round-off error for AffApprox's would be to replace the four complex numbers in the definition of AffApprox by four AComplex numbers complete with their round-off errors, and similarly for the one real number. We will not do this because it would necessitate keeping track of five separate round-off-error terms when we do AffApprox calculations.

Instead, we will replace the four complex numbers by four XComplex's and push all the round-off error into the *.err* term. Thus, the definition of AffApprox remains essentially the same as in Chapter 6. We note that, in doing an AffApprox calculation, our subsidiary calculations will generally be on XComplex numbers and produce an AComplex number whose *.e* term will be plucked off and forced into the *.err* term of the final AffApprox.

Conventions 8.3: i) In what follows, we will use Basic Properties 7.0 and Lemmas 7.0 and 7.1. Also, the Propositions in Chapter 6 will be utilized; as such, the numbering of the Propositions is the same in Chapters 6 and 8 (for example, Proposition 6.7 corresponds to Proposition 8.7).

ii) Some notational simplifications will be introduced: $dist(x)$ before Proposition 8.6, ax before Propositions 8.8, 8.9, 8.11, and ay before Propositions 8.8, 8.9,

iii) We will try to keep our notation fairly consistent with that of the computer programs *verify* and *corona* and this will produce some mildly peculiar notation. In particular, in the operations pertaining to Propositions 8.2 and beyond, the resultant AffApprox's will be denoted $(r\_f.z; r\_f_1.z, r\_f_2.z, r\_f_3.z; r\_error)$ where the first four terms are XComplex's and the last term is a double (technically, $r\_f.z$ is the XComplex part of the AComplex r_f and similarly for the $r\_f_k.z$ terms). One break with the notation of the programs though is that AffApprox's are called (in the programs) ACJ's, which stands for "Approximate Complex 1-Jets."

iv) The propositions that follow include in their statements the definitions of the various operations on AffApprox's (see Remark 6.6iii).

v) We remind the reader (see Section 7.4) that all machine operations act on machine numbers, and that the various variables appearing in the propositions are assumed to be doubles.

-X:

Proposition 8.1: If x is an AffApprox then $S(-x) = -(S(x))$ where

$$-x \equiv (-x.f; -x.f_0, -x.f_1, -x.f_2; x.err)$$

∎

X+Y:

We analyze the addition of the AffApprox's $x = (x.f; x.f_0, x.f_1, x.f_2; x.err)$ and $y = (y.f; y.f_0, y.f_1, y.f_2; y.err)$. To get the first term in $x + y$ we add the XComplex numbers

$x.f$ and $y.f$; which produces the AComplex number $r\_f = x.f + y.f$ (see Proposition 7.4), and then we pluck off the XComplex part, which we denote $r\_f.z$. The round-off error part $r\_f.e$ will be foisted into the overall error term r_error for $x + y$. Similarly for the next three terms in $x + y$.

Abstractly, the overall error term r_error comes from adding the round-off error contributions $r\_f.e,\ r\_f_0.e,\ r\_f_1.e,\ r\_f_2.e$ and the AffApprox error contributions $x.err,\ y.err$. Of course, we have to produce a machine version.

Proposition 8.2: If x and y are AffApprox's, then $S(x + y) \supseteq S(x) + S(y)$, where

$$x + y \equiv (r\_f.z; r\_f_0.z, r\_f_1.z, r\_f_2.z; r\_error)$$

with

$$r\_f = x.f + y.f$$

$$r\_f_k = x.f_k + y.f_k$$

$$r\_error = (1 + 3EPS) \otimes ((x.err \oplus y.err) \oplus ((r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e)))$$

Proof: The error is given by

$$(x.err + y.err) + ((r\_f.e + r\_f_0.e) + (r\_f_1.e + r\_f_2.e))$$

$$\le (1 + EPS/2)(x.err \oplus y.err) + (1 + EPS/2)((r\_f.e \oplus r\_f_0.e) + (r\_f_1.e \oplus r\_f_2.e))$$

$$\le (1 + EPS/2)^3((x.err \oplus y.err) \oplus ((r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e)))$$

$$\le (1 + 3EPS) \otimes ((x.err \oplus y.err) \oplus ((r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e)))$$

To get the last line we used Lemma 7.1. ∎

X - Y:

Proposition 8.3: If x and y are AffApprox's, then $S(x - y) \supseteq S(x) - S(y)$, where

$$x - y \equiv (r\_f.z; r\_f_0.z, r\_f_1.z, r\_f_2.z; r\_error)$$

with

$$r\_f = x.f - y.f$$

$$r\_f_k = x.f_k - y.f_k$$

$$r\_error = (1 + 3EPS) \otimes ((x.err \oplus y.err) \oplus ((r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e)))$$

∎

X + D:

Here, we add the AffApprox $x = (x.f; x.f_0, x.f_1, x.f_2; x.err)$ to the double y. The only terms that change are the first and the last.

Proposition 8.4: If x is an AffApprox and y is a double, then $S(x+y) \supseteq S(x) + S(y)$, where

$$x + y \equiv (r\_f.z; r\_f_0.z, r\_f_1.z, r\_f_2.z; r\_error)$$

with

$$r\_f = x.f + y$$

$$r\_f_k = x.f_k$$

$$r\_error = (1 + EPS) \otimes (x.err \oplus r\_f.e)$$

Proof: The error is given by

$$x.err + r\_f.e$$

$$\leq (1 + EPS) \otimes (x.err \oplus r\_f.e)$$

by Lemma 7.0. ∎

X - D:

Proposition 8.5: If x is an AffApprox and y is a double, then $S(x-y) \supseteq S(x) - S(y)$, where

$$x - y \equiv (r\_f.z; r\_f_0.z, r\_f_1.z, r\_f_2.z; r\_error)$$

with

$$r\_f = x.f - y$$

$$r\_f_k = x.f_k$$

$$r\_error = (1 + EPS) \otimes (x.err \oplus r\_f.e)$$

∎

X × Y :

We multiply the AffApprox's x and y while pushing all error into the $.err$ term.

We will use the functions (see Formulas 7.0 and 7.1, at the end of Chapter 7) $absUB = (1 + 2EPS) \otimes hypot_o(x.re, x.im)$ and $absLB(x) = (1 - 2EPS) \otimes hypot_o(x.re, x.im)$.

When x is an AffApprox, we define $dist(x)$ to be

$$(1 + 2EPS) \otimes (absUB(x.f_0) \oplus (absUB(x.f_1) \oplus absUB(x.f_2))).$$

This is the machine representation of the sum of the absolute values of the linear terms in the AffApprox x (the proof is straightforward).

Proposition 8.6: If x and y are AffApprox's, then $S(x \times y) \supseteq S(x) \times S(y)$, where

$$x \times y \equiv (r\_f.z; r\_f_0.z, r\_f_1.z, r\_f_2.z; r\_error)$$

with

$$r\_f = x.f \times y.f$$

$$r\_f_k = x.f \times y.f_k + x.f_k \times y.f$$

$$r\_error = (1 + 3EPS) \otimes (A \oplus (B \oplus C))$$

and

$$A = (dist(x) \oplus x.err) \otimes (dist(y) \oplus y.err)$$

$$B = absUB(x.f) \otimes y.err \oplus absUB(y.f) \otimes x.err$$

$$C = (r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e)$$

Proof: We add the non-round-off error term for $x \times y$ to the various round-off error terms that accumulated.

$$((dist(x) + x.err) \times (dist(y) + y.err)) + ((absUB(x.f) \times y.err + absUB(y.f) \times x.err) + (r\_f.e + r\_f_0.e) + (r\_f_1.e + r\_f_2.e))$$

$$\leq (1+EPS/2)^3[(dist(x) \oplus x.err) \otimes (dist(y) \oplus y.err)] + (1+EPS/2)^2\{(absUB(x.f) \otimes y.err \oplus absUB(y.f) \otimes x.err) + ((r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e))\}$$

$$\leq (1 + EPS/2)^3 A + (1 + EPS/2)^3 (B \oplus C)$$

$$\leq (1 + 3EPS) \otimes (A \oplus (B \oplus C))$$

∎

X × D :

Proposition 8.7: If x is an AffApprox and y is a double, then $S(x \times y) \supseteq S(x) \times S(y)$, where

$$x \equiv y = (r\_f.z; r\_f_0.z, r\_f_1.z, r\_f_2.z; r\_error)$$

with

$$r\_f = x.f \times y$$

$$r\_f_k = x.f_k \times y$$

$$r\_error = (1 + 3EPS) \otimes ((x.err \otimes |y|) \oplus ((r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e)))$$

∎

X/Y:

For convenience, let $ax = absUB(x.f),\ ay = absLB(y.f)$.

Proposition 8.8: If x and y are AffApprox's with $D > 0$ (see below), then $S(x/y) \supseteq S(x)/S(y)$, where

$$x/y \equiv (r\_f.z; r\_f_0.z, r\_f_1.z, r\_f_2.z; r\_error)$$

with

$$r\_f = x.f/y.f$$

$$r\_f_k = (x.f_k \times y.f - x.f \times y.f_k)/(y.f \times y.f)$$

$$r.error = (1 + 3EPS) \otimes (((1 + 3EPS) \otimes A \ominus (1 - 3EPS) \otimes B) \oplus C)$$

and

$$A = (ax \oplus (dist(x) \oplus x.err)) \oslash D$$

$$B = (ax \oslash ay \oplus dist(x) \oslash ay) \oplus ((dist(y) \otimes ax) \oslash (ay \otimes ay))$$

$$C = (r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e)$$

$$D = ay \ominus (1 + EPS) \otimes (dist(y) \oplus y.err)$$

Proof: As usual, we add the round-off errors to the old AffApprox error, taking into account round-off error. Let's work on it bit by bit.

$$(ax + dist(x) + x.err)/(ay - (dist(y) + y.err))$$

$$\leq (1 + EPS/2)^2 (ax \oplus (dist(x) \oplus x.err))/(ay - (1 + EPS) \otimes (dist(y) \oplus y.err))$$

$$\leq (1 + EPS/2)^2 (ax \oplus (dist(x) \oplus x.err))/(\frac{1}{1 + EPS/2})(ay \ominus (1 + EPS) \otimes (dist(y) \oplus y.err))$$

$$\leq (1 + EPS/2)^4 (ax \oplus (dist(x) \oplus x.err)) \oslash (ay \ominus (1 + EPS) \otimes (dist(y) \oplus y.err))$$

$$\leq (1 + 3EPS) \otimes A$$

The next term, being subtracted, requires opposite inequalities.

$$(ax/ay + dist(x)/ay) + dist(y) \times ax/(ay \times ay)$$

$$\geq (1 - EPS/2)(ax \oslash ay + dist(x) \oslash ay) + (1 - EPS/2)(dist(y) \otimes ax)/(\frac{1}{1 - EPS/2})(ay \otimes ay)$$

$$\geq ((1 - EPS/2)^4 ((ax \oslash ay \oplus dist(x) \oslash ay) \oplus ((dist(y) \otimes ax) \oslash (ay \otimes ay)))$$

$$\geq (1 + EPS/2)(1 + 3EPS)(B)$$

$$\geq (1 - 3EPS) \otimes B$$

Finally, we do the round-off terms.

$$((r\_f.e + r\_f_0.e) + (r\_f_1.e + r\_f_2.e))$$

$$\leq (1 + EPS/2)^2 C$$

Now, we put these three pieces together.

$$(ax + dist(x) + x.err)/(ay - (dist(y) + y.err))$$
$$-((ax/ay + dist(x)/ay) + dist(y) \times ax/(ay \times ay)) + ((r\_f.e + r\_f_0.e) + (r\_f_1.e + r\_f_2.e))$$

$$\leq (1+3EPS)\otimes A - (1-3EPS)\otimes B + (1+EPS/2)^2 C$$

$$\leq (1+EPS/2)^2(((1+3EPS)\otimes A \ominus (1-3EPS)\otimes B) + C)$$

$$\leq (1+3EPS)\otimes(((1+3EPS)\otimes A \ominus (1-3EPS)\otimes B) \oplus C)$$

∎

D/X:

We are dividing a double x by an AffApprox y. For convenience, let $ax = |x|$, $ay = absLB(y.f)$. Having done division out in the previous proposition, we will skip the proof of Proposition 8.9. See the *Annals* archive for the proof.

Proposition 8.9: If x is a double and y is an AffApprox with $D > 0$ (see below), then $S(x/y) \supseteq S(x)/S(y)$, where

$$x/y \equiv (r\_f.z; r\_f_0.z, r\_f_1.z, r\_f_2.z; r\_error)$$

with

$$r\_f = x/y.f$$

$$r\_f_k = -(x \times y.f_k)/(y.f \times y.f)$$

$$r\_error = (1+3EPS)\otimes(((1+2EPS)\otimes(ax \oslash D) \ominus (1-3EPS)\otimes B) \oplus C)$$

$$B = ax \oslash ay \oplus (dist(y) \otimes ax \oslash (ay \otimes ay))$$

$$C = (r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e)$$

$$D = ay \ominus (1+EPS)\otimes(dist(y) \oplus y.err)$$

∎

X/D:

We are dividing an AffApprox x by a double y (the computer will object if $y = 0$). The proof is easy, and so, we delete it.

Proposition 8.10: If x is an AffApprox and y is a double, then $S(x/y) \supseteq S(x)/S(y)$, where

$$x/y \equiv (r\_f.z; r\_f_0.z, r\_f_1.z, r\_f_2.z; r\_error)$$

with

$$r\_f = x.f/y$$

$$r\_f_k = x.f_k/y$$

$$r\_error = (1+3EPS)\otimes((x.err \oslash |y|) \oplus [(r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e)])$$

∎

$\sqrt{\mathbf{X}}$:

Here, x is an AffApprox and we let $ax = absUB(x.f)$. There are two cases to consider depending on whether or not $D = ax \ominus (1+EPS) \otimes (dist(x) \oplus x.err)$ is or is not greater than zero.

Proposition 8.11a: If x is an AffApprox and $D = ax \ominus (1+EPS) \otimes (dist(x) \oplus x.err)$ is not greater than zero, then $S(\sqrt{x}) \supseteq \sqrt{S(x)}$, where we use the crude overestimate

$$\sqrt{x} \equiv (0; 0, 0, 0; (1+2EPS) \otimes \sqrt[\circ]{(ax \oplus (xdist \oplus x.err))}\ )$$

Proof:

$$\sqrt{ax + xdist + x.err}$$
$$\leq (1+EPS/2)\sqrt{(ax \oplus (xdist \oplus x.err))}$$
$$\leq (1+2EPS) \otimes \sqrt[\circ]{(ax \oplus (xdist \oplus x.err))}$$

∎

Proposition 8.11b: If x is an AffApprox and $D = ax \ominus (1+EPS) \otimes (dist(x) \oplus x.err)$ is greater than zero, then $S(\sqrt{x}) \supseteq \sqrt{S(x)}$, where

$$\sqrt{x} \equiv (r\_f.z; r\_f_0.z, r\_f_1.z, r\_f_2.z; r\_error)$$

with

$$r\_f = \sqrt{x.f}$$
$$t = r\_f + r\_f$$
$$r\_f_k = \text{AComplex}(x.f_k.re, x.f_k.im; 0)/t$$

(Simply put, $r\_f_k = x.f_k/(2\sqrt{x.f})$. The reason we have to fuss to define r_f_k is because $\sqrt{x.f}$ is an AComplex.)

$$r\_error = (1+3EPS) \otimes ($$
$$\{(1+EPS) \otimes \sqrt[\circ]{ax}$$
$$\ominus (1-3EPS) \otimes [dist(x) \oslash (2 \times \sqrt[\circ]{ax}) \oplus \sqrt[\circ]{D}\ ]\}$$
$$\oplus ((r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e))$$
$$)$$

Proof: Let's work on the pieces.

$$\sqrt{ax} \leq (1+EPS) \otimes \sqrt[\circ]{ax}$$

Next,

$$dist(x)/(2\sqrt{ax}) + \sqrt{ax - (dist(x) + x.err)}$$

$$\geq (1 - EPS/2)^2 dist(x) \oslash (2\sqrt[\circ]{ax})$$
$$+ (1 - EPS/2)^{1/2}\sqrt{ax \ominus (1 + EPS) \otimes (dist(x) \oplus x.err)}$$
$$\geq (1 - EPS/2)^3[dist(x) \oslash (2\sqrt[\circ]{ax}) \oplus \sqrt[\circ]{D}\ ]$$
$$\geq (1 + EPS/2)(1 - 3EPS)[dist(x) \oslash (2\sqrt[\circ]{ax}) \oplus \sqrt[\circ]{D}\ ]$$
$$\geq (1 - 3EPS) \otimes [dist(x) \oslash (2\sqrt[\circ]{ax}) \oplus \sqrt[\circ]{D}\ ]$$

Adding in the usual term, we get as our error bound

$$\sqrt{ax} - (dist(x)/(2\sqrt{ax}) + \sqrt{ax - (dist(x) + x.err)}\ ) + ((r\_f.e + r\_f_0.e) + (r\_f_1.e + r\_f_2.e))$$

$$\leq (1 + EPS) \otimes \sqrt[\circ]{ax}$$
$$- (1 - 3EPS) \otimes [dist(x) \oslash (2\sqrt[\circ]{ax}) \oplus \sqrt[\circ]{D}\ ]$$
$$+ (1 + EPS/2)^2((r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e))$$

$$\leq (1 + EPS/2)^3(\{(1 + EPS) \otimes \sqrt[\circ]{ax}$$
$$\ominus (1 - 3EPS) \otimes [dist(x) \oslash (2\sqrt[\circ]{ax}) \oplus \sqrt[\circ]{D}\ ]\}$$
$$\oplus ((r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e)))$$

$$\leq (1 + 3EPS) \otimes (\{(1 + EPS) \otimes \sqrt[\circ]{ax}$$
$$\ominus (1 - 3EPS) \otimes [dist(x) \oslash (2\sqrt[\circ]{ax})$$
$$\oplus \sqrt[\circ]{D}\ ]\} \oplus ((r\_f.e \oplus r\_f_0.e) \oplus (r\_f_1.e \oplus r\_f_2.e)))$$

∎

Appendix I: Review of the Theory of Insulators in Hyperbolic 3-Manifolds

Given that Theorem 0.2 is the main technical result of this paper, we herewith present an appendix which briefly recalls from [G] various definitions, examples and properties of insulators in hyperbolic 3-manifolds. It furthermore briefly recalls why the existence of a non-coalescable insulator family for a geodesic δ in a closed orientable hyperbolic 3-manifold N allows us to conclude that any irreducible 3-manifold M homotopy equivalent to N is indeed homeomorphic to N.

Definition A.1: Let G be a group of homeomorphisms of S^2 and $\mathcal{A} = \{A_i\}$ a countable set of pairwise disjoint G-equivariant pairs of points of S^2, i.e. if $g \in G, A_i \in \mathcal{A}$, then $g(A_i) \in \mathcal{A}$. Let $\{\lambda_{ij}\}$ be a collection of smooth simple closed curves in S^2. $\{\lambda_{ij}\}$ is called a $(G, \mathcal{A})$ *insulator family* and each λ_{ij} is an *insulator* if

i) *Separation*: If $i \neq j$, then λ_{ij} separates A_i from A_j.

ii) *Equivariance*: If $g \in G$, then $g(\lambda_{ij})$ is the curve associated to the pair $g(A_i), g(A_j)$. Also $\lambda_{ij} = \lambda_{ji}$.

iii) *Convexity*: To each λ_{ij} there exist round circles respectively containing A_i and A_j and disjoint from λ_{ij}.

iv) *Local Finiteness*: For every $\epsilon > 0$ there exist only finitely many λ_{ij} such that i is fixed and $\mathrm{diam}(\lambda_{ij}) > \epsilon$.

Definition A.2: A $(G, \mathcal{A})$ insulator family is *noncoalescable* if it satisfies the following *no trilinking* property. For no i, does there exist $\lambda_{ij_1}, \lambda_{ij_2}, \lambda_{ij_3}$ whose union separates the points of A_i. A hyperbolic 3-manifold satisfies the *insulator condition* if there exists a geodesic δ in N and a $(\pi_1(N), \{\partial\delta_i\})$ noncoalescable insulator family. Here $\{\delta_i\}$ is the set of lifts of δ to $\mathbf{H}^3$.

Example A.3: Let δ be a simple closed geodesic in the hyperbolic 3-manifold N. δ lifts to a collection $\Delta = \{\delta_i\}$ of hyperbolic lines in $\mathbf{H}^3$. To each pair δ_i, δ_j, there exists the *midplane* D_{ij}, i.e. the hyperbolic halfplane orthogonal to and cutting the middle of the *orthocurve* (i.e. the shortest line segment) between δ_i and δ_j. Each D_{ij} extends to a circle λ_{ij} on S^2_∞, which separates $\partial\delta_i$ from $\partial\delta_j$. We call the insulator family $\{\lambda_{ij}\}$ the *Dirichlet insulator family.* If tuberadius$(\delta) > \log(3)/2$, then this family is noncoalescable, the idea being that from the point of view of δ each λ_{ij} takes up less than 120 degrees of visual angle, thus there can be no trilinkings among the λ_{ij}'s.

The content of Chapter 2 is a construction of a new insulator family for δ called the Corona insulator family. If tuberadius$(\delta) > \log(3)/2$, then this family is just the Dirichlet family. Lemma 2.5 provides a sufficient condition for the Corona family to be noncoalescable.

Theorem 0.2 is established by showing that the Corona family for δ is non-coalescable if $N \neq$ Vol3 while if $N =$ Vol3, then an insulator which is a hybrid of the Dirichlet and Corona insulators is non-coalescable.

Non-coalescable insulator families are important because of the following result.

Theorem A.4 [G]: Let N be a closed orientable hyperbolic 3-manifold containing a geodesic δ having a non-coalescable insulator family, then

i) If $f : M \to N$ is a homotopy equivalence where M is an irreducible 3-manifold, then f is homotopic to a homeomorphism.

ii) If $f, g : N \to N$ are homotopic homeomorphisms, then f is isotopic to g.

iii) The space of hyperbolic metrics on N is path connected.

Idea of the proof of i) The plan is to find a simple closed curve γ in M and a homotopy of f to $g : M \to N$ so that $g|N(\gamma) : N(\gamma) \to N(\delta)$ is a homeomorphism, $g(M - \mathring{N}(\gamma)) = N - \mathring{N}(\delta)$ and $g|M - \mathring{N}(\gamma) : M - \mathring{N}(\gamma) \to N - \mathring{N}(\delta)$ is a homotopy equivalence. Since $N - \mathring{N}(\delta)$ is Haken we invoke Waldhausen [Wa] to conclude that $g|M - \mathring{N}(\gamma)$ is homotopic to a homeomorphism rel boundary and hence f is homotopic to a homeomorphism.

There are many technical difficulties in carrying out the above plan, the primary one being how to find the curve γ. Here is another view of this problem. Given a hyperbolic 3-manifold Q, it is well known that any nontrivial element α of $\pi_1(Q)$ is represented by a unique geodesic η in Q. Suppose that we are given Q with some random Riemannian metric and we do not know the hyperbolic metric, then how are we supposed to find η? In our setting if $f_* : \pi_1(M) \to \pi_1(N)$ is the induced map on fundamental group then our goal is to find the γ which represents $f_*^{-1}([\delta])$. We outline how to solve this problem if δ has a noncoalescable insulator family.

Theorem A.5 [G]: If $f : M \to N$ is a homotopy equivalence, where M is an irreducible 3-manifold and N is a closed hyperbolic 3-manifold, then there exists a closed hyperbolic 3-manifold X and regular covering maps $p_1 : X \to M$, $q_1 : X \to N$ such that $f \circ p_1$ is homotopic to q_1. A lift $\tilde{f} : \mathbf{H}^3 \to \mathbf{H}^3$ extends to id:$S^2_\infty \to S^2_\infty$. Furthermore the action of $\pi_1(M)$ on $\mathbf{H}^3$ extends to a Mobius action on S^2_∞ which is identical to the action of $\pi_1(N)$ on S^2_∞.

In words this theorem says that $\tilde{M}$ and $\tilde{N}$ have common spheres at infinity. Thus λ_{ij} is a $\{\pi_1(M), \{\partial\delta_i\}\}$ non-coalescable insulator family.

Here is how to construct γ. To each smooth simple closed curve λ_{ij} in S^2_∞, there exists a lamination $\sigma_{ij} \subset \mathbf{H}^3$ by embedded least area planes, with limit set λ_{ij} such that σ_{ij} lies in a uniformly fixed width hyperbolic regular neighborhood of the hyperbolic convex hull of λ_{ij}. Here $\mathbf{H}^3$ is given the Riemannian metric induced from M, while the fixed width measurement is via the hyperbolic metric induced from N. Fix i. Let H_{ij} be the $\mathbf{H}^3$-complementary region of σ_{ij} containing the ends of δ_i. It turns out that $\cap_j H_{ij}$ contains a component $V_i = (\mathring{D})^2 \times \mathbf{R}$ which projects to an open solid torus in M. Define γ to be the core of this solid torus and γ_i the lift which lives in V_i. Up to isotopy, γ is independent of all choices, in particular the choice of Riemannian metric on M.

To first approximation think of σ_{ij} as a properly embedded plane and V_i as the intersection of topological half spaces in $\mathbf{H}^3$. (It is not known if the leaves of σ_{ij} must be properly embedded.) Without the no-trilinking condition this intersection might be empty. Note that if $\mathbf{H}^3$ is given the hyperbolic metric and λ_{ij} is the Dirichlet insulator, then each σ_{ij} is a totally geodesic plane and the projection of V_i into N is a solid torus regular neighborhood of δ. This statement is also true if λ_{ij} is any non-coalescable insulator family, however the proof requires the convexity condition. Thus the above construction using the Riemannian metric induced from M (resp. N) yields the curves $\{\gamma_i\} \subset \mathbf{H}^3$ (resp. $\{\delta_i\}$). These collections $\{\gamma_i\} = \Gamma$ and $\{\delta_i\} = \Delta$ are $\mathbf{B}^3$-links (i.e. sets of pairwise disjoint properly embedded arcs in $\mathbf{B}^3$ whose restriction to $\mathbf{H}^3$ is locally finite).

We now give a hint as to how to find the desired $g : M \to N$. The Riemannian metric μ_0 on X induced from M and the hyperbolic metric μ_1 on X induced from N lift to $\pi_1(X)$ equivariant metrics $\tilde{\mu}_t, t \in \{0, 1\}$ on $\mathbf{H}^3$, so the above construction applied to the $(\pi_1(X), \{\partial\delta_i\})$ insulator family $\{\lambda_{ij}\}$ with respect to the $\tilde{\mu}_t$ metric yields a link τ_t in X. Since the isotopy class of τ_t is independent of t, $\tau_0 = p_1^{-1}(\gamma)$ is isotopic to $\tau_1 = q_1^{-1}(\delta)$. We conclude that the $\mathbf{B}^3$-link Γ is equivalent to the $\mathbf{B}^3$-link Δ, i.e. there exists a homeomorphism $k : (\mathbf{B}^3, \Gamma) \to (\mathbf{B}^3, \Delta)$ so that $k \mid S^2_\infty = \mathrm{id}$.

The above mentioned g arises in the midst of proving the next result. Here $p : \mathbf{H}^3 \to M$, and $q : \mathbf{H}^3 \to N$ are the universal covering maps arising from A.5.

Proposition A.6 [G]: Let $f : M \to N$ be a homotopy equivalence between the closed orientable hyperbolic 3-manifold N and the irreducible manifold M. If there exists a simple closed curve $\gamma \subset M$, a geodesic $\delta \subset N$ and a homeomorphism $k : (\mathbf{B}^3, p^{-1}(\gamma)) \to (\mathbf{B}^3, q^{-1}(\delta))$ such that $k \mid \partial\mathbf{B}^3 = \mathrm{id}$, then f is homotopic to a homeomorphism.

Remark A.7: As mentioned above, if ρ_0 is a hyperbolic metric on N, then associated to a nontrivial element α of $\pi_1(N)$, there exists a unique geodesic $\delta_0 \subset N$ representing α. A second hyperbolic metric ρ_1 on N will give rise to the geodesic δ_1. The Mostow Rigidity theorem implies that there exists an isometry of N_{ρ_0} to N_{ρ_1} homotopic to the id. A consequence of Theorem 0.1 is that the isometry is *isotopic* to the identity. Thus Theorem 0.1 is a strengthening of Mostow Rigidity. To appreciate the difference, note that the Mostow Rigidity theorem does not imply that δ_0 is isotopic to δ_1, while Theorem 0.1 does. The proof is a consequence of the insulator technology. Here is a hint. The construction of γ is independent, up to isotopy, of all choices. Thus applying the insulator construction to N respectively using the metrics ρ_0 and ρ_1 yields the curves δ_0 and δ_1, which are necessarily isotopic.

References

[Beardon] A. Beardon, *The Geometry of Discrete Groups*, Graduate Texts in Mathematics 91, Springer-Verlag, New York, 1983.

[Be] J. Berge, *Heegaard*, A program to understand Heegaard splittings of 3-manifolds, available free from the author.

[CHW] P. Callahan, M. Hildebrand, and J. Weeks, A Census of Cusped Hyperbolic 3-Manifolds, *Math. Comp.* **68** (1999), 321–332.

[CGHN], D. Coulson, O. Goodman, C. Hodgson, and W. Neumann, Computing Arithmetic Invariants of 3-Manifolds, to appear in *Exp. Math.*, March 2000.

[CS] M. Culler and P. Shalen, Varieties of Group Representations and Splittings of 3-Manifolds, it Annals of Math. **117** (1983), 109–146.

[F] W. Fenchel, Elementary Geometry in Hyperbolic Space, de Gruyter Studies Math., 11, de Gruyter, Berlin (1989).

[FS] L. H. de Figueiredo and J. Stolfi, Self-Validated Numerical Methods and Applications, Brazilian Mathematics Colloquium monograph, IMPA, Rio de Janeiro, Brazil (1997).

[G] D. Gabai, On the Geometric and Topological Rigidity of Hyperbolic 3-Manifolds, *J. Amer. Math. Soc.* **10** (1997), 37–74.

[GM1] F. Gehring and G. Martin, Inequalities of Möbius Transformations and Discrete Groups, *J. Reine Agnew. Math.* **418** (1991), 31–76.

[GM2] F. Gehring and G. Martin, Precisely Invariant Collars and the Volume of Hyperbolic 3-folds, *J. Differential Geom.* **49**(1998), 411–435.

[HW1] C. Hodgson and J. Weeks, personal communication.

[HW2] C. Hodgson and J. Weeks, Symmetries, Isometries and Length Spectra of Closed Hyperbolic 3-Manifolds, *Exp. Math.* **3** (1994), 261–274.

[IEEE] IEEE Standard for Binary Floating-Point Arithmetic (ANSI/IEEE Std 754-1985) published by the Institute of Electrical and Electronics Engineers, Inc., NY, NY, 1985.

[JR] K. Jones and A. Reid, Vol3 and Other Exceptional Hyperbolic 3-Manifolds, to appear *Proc. AMS*.

[K1] W. Kahan, Interval Arithmetic Options in the Proposed IEEE Floating Point Arithmetic Standard; *Interval Mathematics 1980*, Ed. Karl L. E. Nickel, 99–128.

[K2] W. Kahan, Beastly Numbers, preprint.

La] O. Lanford, Computer-Assisted Proofs in Analysis, in *Proceedings of the ICM*, Berkeley, California, 1986, vol. 2, 1385–1394.

[Math] Mathematica 2.0, A system for doing mathematics by computer, Wolfram Research Inc., Champaign Il.

[M1] G. R. Meyerhoff, A Lower Bound for the Volume of Hyperbolic 3-Manifolds, *Canadian J. Math.* **39** (1987), 1038–1056.

[M2] G. R. Meyerhoff, Sphere-Packing and Volume in Hyperbolic 3-Space, *Comment. Math. Helvetici* **61** (1986), 271-278.

[Mo] G. D. Mostow, Quasi-conformal Mappings in n-Space and the Rigidity of Hyperbolic Space Forms, *IHES Publ. Math.* **34** (1968), 53-104.

[Pr] A. Przeworski, Tubes in Hyperbolic 3-Manifolds, preprint.

[Ri] R. Riley, Applications of a Computer Implementation of Poincaré's Theorem on Fundamental Polyhedra, *Math. of Computation* **40** (1983) 607–632.

[T] N. Thurston, Finding Killer Words, in preparation.

[Wa] F. Waldhausen, On Irreducible 3-Manifolds which Are Sufficiently Large, *Annals of Math.* **87** (1968), 56-88.

[W1] J. Weeks, *SnapPea*, available free from the author.

[W2] J. Weeks, personal communication.