ORIGINAL

Sprint

Charles J. Rehwinkel
General Attorney

PO Box 2214
Tallahassee, FL 32316
Mailstop FLTLH00107
Voice 850 847 0244
Fax 850 599 1458

September 2, 1998

Ms. Blanca S. Bayo, Director
Division of Records and Reporting
Florida Public Service Commission
2540 Shumard Oak Boulevard
Tallahassee, Florida 32399-0850

Re: Docket No. 980696-TP

Dear Ms. Bayo:

Enclosed for filing in the above docket are the original and fifteen (15) copies of the Direct Testimonies of Carl H. Laemmli, Kent W. Dickerson, Brian K. Staihr and James W. Sichter on behalf of Sprint-Florida, Incorporated.

Please acknowledge receipt and filing of the above by stamping the duplicate copy of this letter and returning the same to this writer.

Thank you for your assistance in this matter.

Sincerely,

Charles J. Rehwinkel

CJR/th

Enclosures

ACK ___
AFA 2
APP ___
CAF ___
CMU ___
CTR ___
EAG ___
LEG 2
LIN 5 + org
OPC ___
RCH ___
SEC 1
WAS ___
OTH ___

RECEIVED & FILED
FPSC-BUREAU OF RECORDS

Laemmli
DOCUMENT NUMBER-DATE
09589 SEP-2 98
FPSC-RECORDS/REPORTING

Dickerson
DOCUMENT NUMBER-DATE
09590 SEP-2 98
FPSC-RECORDS/REPORTING

Staihr
DOCUMENT NUMBER-DATE
09591 SEP-2 98
FPSC-RECORDS/REPORTING

Sichter
DOCUMENT NUMBER-DATE
09592 SEP-2 98
FPSC-RECORDS/REPORTING

ORIGINAL

BEFORE THE FLORIDA PUBLIC SERVICE COMMISSION

REBUTTAL TESTIMONY OF BRIAN K. STAIHR

ON BEHALF OF SPRINT-FLORIDA. INCORPORATED

DOCKET 980696-TP

SEPTEMBER 2, 1998

Q. Please state your name, title and business address.

A. My name is Brian K. Staihr. I am employed by Sprint United Management Company ("Sprint") as Regulatory Economist. My business address is 4220 Shawnee Mission Parkway, Suite 303, Fairway, KS, 66205.

Q. Are you the same Brian Staihr who filed direct testimony in this proceeding on August 3, 1998?

A. Yes I am.

Q. What is the purpose of your rebuttal testimony?

A. In my rebuttal testimony I address specific points raised by Don Wood with regard to the HAI Model Version 5.0a, filed in this proceeding by MCI and AT&T. I also address certain comments made by Mr. Wood regarding the Benchmark Cost Proxy Model (BCPM) Version 3.1 as filed by Sprint.

DOCUMENT NUMBER-DATE
09591 SEP -2 98
FPSC-RECORDS/REPORTING

Q. Please provide a summary statement of your rebuttal testimony.

A. Contrary to Mr. Wood's statements, the HAI Model is not "the most accurate and reliable means" of developing cost information (Wood Direct p. 3). In the past several months, significant problems have been identified at both the Federal and state levels regarding the accuracy of the HAI Model 5.0a, as filed in this proceeding. These problems have specifically involved the HAI Model's preprocessing, customer location algorithms, network construction algorithms, and various assumptions built into the model and the model sponsors' values for certain user-adjustable inputs. Several of these problems are discussed in the testimony that follows.

Q. In his testimony Mr. Wood lists two states, Kentucky and Louisiana, where the Commissions chose to rely on the HAI Model for USF purposes. Do these two decisions provide evidence that the problems you mention above were of no concern to these Commissions?

A. Absolutely not. It is important to understand that a great deal of information and analysis regarding the HAI Model has come to light only in the past few months. This is because a large portion of the information used by the HAI Model in its preprocessing stages was only made available to parties (under order of the Nevada Commission) in April of this year. I specifically refer to the geocoded locations that are placed within main and outlier clusters. These clusters are then used by the HAI Model.

In April and May of this year Sprint examined this previously unavailable information used by the HAI Model and made several *ex parte* presentations to the FCC. These *ex parte* presentations outlined how this information is used in the HAI Model preprocessing and customer location algorithms (Sprint FCC *ex parte*(s), April 17-30, 1998). *These documents demonstrated that in rural areas the HAI Model systematically underestimates the dispersion of customers. As a result of this underestimation, the HAI produces less distribution plant than the amount that would actually be needed to provide service to the customer locations the model purports to use.*[1] [Footnotes are included as endnotes in Exhibit BKS-1A] All of these *ex parte* documents are on record at the FCC, and I have included several of the documents here as Exhibit BKS-2.

Q. Did the Kentucky and Louisiana Commissions have access to these *ex parte* presentation documents?

A. No. The Louisiana proceeding to which Mr. Wood refers took place in late January (1/28-30). The Kentucky proceeding to which Mr. Wood refers took place at the beginning of March (3/3-6).

Q. At the time of the Kentucky and Louisiana proceedings, did any party to those proceedings have access to the information that served as the basis for those *ex parte* presentations?

A. No. The information that served at the basis for those documents is housed at the economic research firm of PNR & Associates in Jenkintown, Pennsylvania. Until April 15-17, 1998, no party other than PNR had been allowed access to that information.

Q. How did the FCC respond to these *ex parte* presentations?

A. Following these presentations the FCC produced its own analysis of the HAI customer location algorithm conducted by Jeffrey Prisbrey. This analysis and Sprint's response to it are attached as Exhibit BKS-3. The results of Mr. Prisbrey's analysis support Sprint's findings: That the HAI Model method "underestimates the dispersion" of customer locations (Prisbrey page 3). According to Prisbrey, this underestimation is most extreme when clusters consist of small numbers of customers, as is often the case in rural areas. This underestimation causes the model to build insufficient plant, because it builds to locations that are closer together than the customers' actual locations.

Q. Can you comment on how this information was received, or the impact this information had, in any other state proceedings?

A. In Nevada, Costing Docket # 96-9035, the Nevada Commission initially chose the HAI Model's immediate predecessor, the Hatfield Model 3.1, to be used for unbundled element (UNE) costing with the intent to also use the model for universal service purposes (USF) [Nevada PUC Opinion and Order, March 5 1998]. When it was pointed out in the proceeding that the FCC had rejected the Hatfield Model 3.0, the Commission moved toward the HAI Model 5.0a, again with the intent of using the model for both UNEs and USF [ibid.]. Sprint then filed a report with the Nevada Commission discussing the HAI Model's customer location algorithm and the FCC analysis discussed above [April 22, 1998]. In a subsequent order, the Nevada Commission declined to submit the HAI model to the FCC to be used in calculating universal service support [Docket 97-5018, Nevada PUC Order, May 14, 1998].

In Minnesota, although the Minnesota Public Utilities Commission appears to have adopted the HAI Model for interconnection and UNE issues, the presiding Administrative Law Judge issued questions about the model on July 16, 1998 directly related to this underbuilding issue. Specifically, the ALJ has asked whether the distribution plant constructed within each cluster should be extended further (increased) in order to come closer to the actual amount needed to provide service to purported customer locations. [State of Minnesota, MPUC Docket No. P-442, 5321, 3157, 466, 421/CI-96-1540].

More recently, in the state of Washington, the Washington UTC issued a bench request asking both model sponsors to make adjustments to their models. Specifically in the case of the HAI Model, the Commission asked the HAI Sponsors to make corrections that would address the issues raised in the aforementioned Prisbrey/FCC analysis regarding customer dispersion. [Washington UTC, Universal Service Docket #UT-98031(a), August 26, 1998].

Q. How does this information apply to Mr. Wood's testimony, specifically the cites on pages 6 and 7 from the decisions of the Kentucky and Louisiana Commissions?

A. The cites from both Commission decisions refer to the HAI "locating customers" (Wood Direct page 6). As stated in my direct testimony, there is no question that location is a key driver of cost. However, it is not enough for a model to "locate" customers, because a model must also *use* that location information when building the network and calculating costs. If a model "locates" customers but then fails to use that information, there is no advantage to locating customers. The *ex parte* presentations attached demonstrate how the HAI Model's preprocessing ignores actual customer locations when it constructs a network in rural areas. The result, particularly in rural areas, is an understating of the cable required to serve customers. Hence, the HAI Model is not the "most accurate and reliable means" of cost estimation for USF purposes.

Q. Does the HAI Model use geocoded customer location information when it constructs its network?

A. No, it does not. Geocoded locations are only used in the model's preprocessing to determine *which* customers will be served together. Once that has been determined, geocoded location information is never again used.[2] That is why the HAI model produces less plant than is actually required to serve customers.

Q. Since the HAI Model does not build to actual locations, is there a significant advantage to using geocoded information just to determine which customers will be served together, as is done in the HAI Model?

A. Not really. The BCPM considered using geocoded data and rejected the idea for two specific reasons.

First, it is important to realize that geocoding is far from an exact science. The latitude/longitude coordinates assigned to any given street address can vary significantly from geocoder to geocoder, especially in rural areas. A simple example of this is shown in Exhibit BKS-4. On this sheet we have six actual Florida street addresses that have been geocoded by two separate systems. As you can see, each of the systems has placed the customers in a very different location, despite the fact that each system classifies this point as a "street address", the finest level of geocoding available. According to the HAI Model, each of these addresses is an exact location. The question that remains, however, is, which of these exact locations is right?

Second, in rural areas (the areas of most concern for universal service purposes) that street-address level data generally does not exist, and the data that does exist is often of questionable quality. Sprint recently filed comments at the FCC that explain how the use of *some* geocoded data in a cost model can often be worse than using none at all. A copy of these comments is attached as Exhibit BKS-5. Nonetheless, the BCPM is *capable* of using geocoded data, as requested by the FCC, to assign customers to areas which would be grouped together to form serving areas, much in the same way the HAI Model groups customers. For this proceeding Sprint undertook an analysis to determine exactly how much difference it would make to use geocoded data. The result of the analysis showed that it makes very little difference.

Q. Please describe that analysis.

A. As I stated earlier, the only way the HAI Model uses actual customer locations is to determine which customers will be served in which cluster. The BCPM builds to areas called *grids*, not clusters, and customers are assigned to grids through a detailed algorithm described in the BCPM Model Methodology. Since all grids are based on latitude and longitude, it is a straightforward process to use latitude/longitude coordinates of geocoded points to assign customers to grids and proceed from there. This of course assumes that good latitude/longitude data exists.

In this analysis we took 3 specific wire centers from Sprint's operating territory in Florida for which we had reasonably good geocoded data. The 3 wire centers were Inverness, Beverly Hills and Avon Park. The total number of lines served by these 3 wire centers is slightly over 50,000. Using actual customer locations, we assigned residences and businesses to micrsame grids. From that point, micrognids were aggregated into ultimate grids using the standard approach, and the model was re-run. In some cases the new ultimate grids differed from the original ultimate grids because the new placement of customer locations caused the microgrids to be aggregated differently. In other cases, the grids may have remained the same but the actual customer counts and dispersion of customers within the grid may have changed. Our goal was to determine what *costs* and *cable distances* the BCPM would produce using the geocoded locations, and how these costs and distances would compare with the standard BCPM results. These results are shown in the table below. I have attached a more detailed explanation of the geocoding and placement process as Exhibit BKS-6.

[redacted]

Inverness	$39.42	$39.80	0.9%
Beverly Hills	$37.00	$37.53	1.4%
Avon Park	$40.92	$41.51	1.4%
	[redacted]	[redacted]	[redacted]
Inverness	7,361,177	7,391,367	1.8%
Beverly Hills	3,009,300	3,088,937	2.6%
Avon Park	3,091,569	3,207,724	3.8%

As the table shows, the average costs per line vary by less than 1.5% in every case.

More importantly, the amount of network that is built (in terms of route distance) does not vary significantly in the two versions of the model. In every case, the variation was less than 4 percent.

Q. How do you interpret these results?

A. These results provide strong evidence that the original customer location algorithms used in the BCPM are accurate and reliable in providing a standardized way of modeling customer location. In numerous proceedings (including this proceeding, see Wood Direct p. 8) the HAI Sponsors have made the unsupported claim that the BCPM method of placing customers in microgrids based on road mileage was flawed, and inferior to the use of geocoded data. What Mr. Wood does not mention is that in the universal service areas of Florida, 1) the vast majority of the HAI locations are not geocoded and 2) in cases where there is data, the geocoded locations are never used to construct the network anyway!

These results, although clearly a sample, demonstrate that the BCPM approach of initially allocating customers along road miles is valid (which the BCPM Sponsors have always known, based on statistical tests of correlation between road miles and population). Most importantly, they support the conclusion that *without using geocoding* the BCPM is superior to the HAI Model in terms of minimizing the distortion that can occur when one models customer location in rural areas incorrectly.

Q. Specifically, how does this distortion occur in the HAI Model?

A. Once the HAI Model has determined that a certain number of customers will be served in a specific cluster, there is no attempt to maintain the spatial relationship between the customers. The model will distribute the customers' lots uniformly across the area of the cluster. An example of this is shown in Exhibit BKS-7.

In this Exhibit, the dots represent actual customer locations that the HAI purports to use. Panels A, B and C are depictions of various dispersions of eight customer locations. These would be considered the "actual" or geocoded locations. Panel D is a depiction of how the HAI Model will place the eight locations in Panels A, B and C before it builds the network. The exhibit shows how the HAI Model will model the customer locations the same way in every case, despite the fact that the customers are actually situated very differently. Existing distances between customers are ignored, distances which can often be several miles. Also, because the BCPM separates its serving areas into quadrants, the distortion that occurs in Panel A cannot occur in the BCPM. In Panel A, the majority of customers are located in the NW quadrant of the area and none are located in the SW quadrant. In the BCPM, this relationship is maintained: the SW quadrant would contain no customers, and the NW quadrant would contain the number you see in Panel B. In the HAI Model, this does not occur.

Q. You said that once the geocoded data is discarded and the HAI model builds its network, the result of the distortion pictured above is an understatement of cable requirements? Is there evidence of such an underbuilding in the results produced by the HAI Model in this proceeding?

A. Yes there is. Sprint has conducted an analysis for its Florida territories similar to analyses shown in the *ex parte* presentations mentioned above. The results of the Florida analysis are completely consistent with our findings in other states. In the rural areas of Florida, the network "built" by the HAI Model is a non-functioning network. The HAI Model systematically and significantly underbuilds the distribution network.

Q. Please describe how you determined that the HAI underbuilds.

A. The concept is very simple. We examine the amount of network plant that the HAI Model builds within its main clusters.[3] This includes everything on the customer side of the digital loop carrier: the distribution cable, connecting cable[4], and drop cable. All of these are used in the model to do two things: to connect customers to the network (at the DLC) and, by default, to connect customers to each other.

We then examine the distance between the original customer locations as they are used in the HAI Model's preprocessing. This equates to the distances between the blue dots in Exhibit BKS-6, Panels A, B and C. The distance measure used is a *minimum spanning tree* (MST). The minimum spanning tree measures the linear distance required to connect any set of points or customer locations in the most direct way. The length of the MST is what we have determined to be "sufficient". (A minimum spanning tree is discussed and pictured in Exhibit BKS-8.)

In reality, the distance of the MST is usually less than what would be "sufficient" to connect all customers to the network and to each other. The distance of the actual telephone network between a given set of locations (points) is usually longer than the length of the MST for that same set of points. Some reasons for this are: 1) the telephone network usually follows roads (which the MST does not), 2) the telephone network must go up and down hills (the MST assumes the world is flat), and 3) the telephone network must take into account natural barriers such as mountains, lakes, etc. (which the MST ignores.)

However, for our analysis we have assumed that the length of the MST is sufficient. We then compare the length of what the HAI builds to the length of the MST. If the total distance of connecting, distribution and drop cable in a cluster is at least as long as the MST for the points in that cluster, we determine that the Model has not underbuilt that cluster. If the total distance of connecting, distribution and drop cable is *less* than the MST for the points in that cluster, we determine that the HAI has underbuilt that cluster. If a cluster is underbuilt, the network the HAI builds to serve that cluster is non-functioning.

Q. Please describe your findings.

A. In the overwhelming majority of cases the HAI underbuilds the main clusters in rural, low-density areas. As the table below shows, in the lowest density zone the HAI underbuilds over 90% of the main clusters in Sprint's serving territory.

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sprint-United	**0 to 5**	**186**	**169**	**90.8%**
Sprint-Centel	**0 to 5**	**87**	**82**	**94.2%**
Sprint-United	**5 to 20**	**184**	**126**	**68.5%**
Sprint-Centel	**5 to 20**	**214**	**174**	**81.3%**
Sprint-United	**20 to 100**	**314**	**111**	**35.4%**
Sprint-Centel	**20 to 100**	**98**	**38**	**38.8%**

In the table I have separated the next-lowest density zone (5 to 100 lines per square mile) into two parts: 5 to 20 lines per square mile, and 20 to 100 lines per square mile. This split does not exist in either model, but it is valuable as a tool for viewing that this underbuilding problem occurs most frequently in the very low density areas, the exact areas that are of most concern for universal service purposes.

Q. Have other parties used the concept of a minimum spanning tree (MST) as a measure of sufficiency in terms of length?

A. Yes. In the attached FCC analysis, Jeffrey Prisbrey used the same concept to measure customer dispersion. More recently, the FCC staff has been working on a synthesis of the two models presented in this proceeding. This synthesis, termed the HCPM (Hybrid Cost Proxy Model), uses a minimum spanning tree as a measure of sufficiency for outside plant and the algorithm is built into the loop portion of their model.

Q. In other proceedings, have the HAI Sponsors commented on the use of the MST as a measure of "sufficient" plant?

A. Yes they have. Recently in Texas, Dr. Robert Mercer, author of the HAI Model, and Mr. John Klick stated that the MST was an "inappropriate standard" to use in such a comparison. [Supplemental Reply Testimony of Dr. Robert Mercer and Mr. John Klick, Texas PUC Docket #18515, June 10, 1998]. Mercer/Klick went on to state that the "Steiner tree, not the MST, constitutes the minimum true distance required to connect a series of points in a network." [Additional Reply Testimony, Mercer/Klick, Texas PUC Docket #18515, June 30, 1998].

Q. What is a Steiner tree?

A. A Steiner tree is another distance construct from mapping theory. Like the MST it measures distance between a set of points, locations or nodes. However, in the Steiner tree it is possible to add points or nodes in the process of connecting the original points. This can result in an overall shorter distance between points, shorter than the MST. Two simple examples are shown in Exhibit BKS-9, and the concept is discussed in Exhibit BKS-8.

Q. The Mercer/Klick testimony implies that it would be more appropriate to define "sufficient" cable as a distance equal to the Steiner tree, not the minimum spanning tree, in Sprint's analysis. Do you agree?

A. No, I do not. As stated above, in the overwhelming majority of cases the MST distance would actually represent an *insufficient* amount of cable, since it does not account for barriers and constraints that a real-world network must consider. Obviously something less than the MST distance, such as a Steiner tree distance, would be insufficient as well.

Just as importantly, the addition of nodes can only decrease the "required" amount of cable for very few, specific configurations of points. Most of these configurations involve less than five (5) points or locations. It is common knowledge that all HAI main clusters must contain at least five customer locations and most contain many more, even in rural areas. Therefore it is simply incorrect to assume that 1) the Steiner Tree distance will be something shorter than the MST distance, and 2) that the Steiner Tree distance is the appropriate measure of what is "sufficient".

But in the spirit of cooperation Sprint has also conducted an analysis using an equivalent of the Steiner tree. As I state above, it has been shown mathematically [Prim, Exhibit BKS-8] that by adding points or nodes, such as a Steiner tree does, it is sometimes possible in special cases to connect a series of points with less than the MST. But it has been shown that this *reduction* in distance can never be more than 13%. In other words, assume there are 5 households in a HAI main cluster, and the MST tells us it requires 1000 feet of cable to connect them all to each other and to the network. Adding points of interconnection, as the Steiner tree does, might reduce that required amount of cable but it will never reduce it below 870 feet.

In the table below, we present the number of HAI Main clusters in low-density regions that underbuild the network using the Steiner tree as a measure of "sufficient" cable length. The length of the Steiner tree is represented as 87% of the length of the MST.

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sprint-United	**0 to 5**	**186**	**157**	**84.4%**
Sprint-Centel	**0 to 5**	**87**	**80**	**91.9%**
Sprint-United	**5 to 20**	**184**	**109**	**59.2%**
Sprint-Centel	**5 to 20**	**214**	**152**	**71.0%**
Sprint-United	**20 to 100**	**314**	**81**	**25.9%**
Sprint-Centel	**20 to 100**	**98**	**28**	**28.6%**

As the table shows, using the Steiner tree as a measure of "sufficient" cable has little impact. In the overwhelming majority of cases that represent universal service areas, the HAI still underbuilds the network.

Q. The figures above speak to the frequency with which the HAI underbuilds, but what is the magnitude of this shortage?

A. The two tables that follow demonstrate that the magnitude of this shortage is significant. In the first table, I have shown a sample of main clusters from Sprint's serving territory in Florida. The table lists the wire center associated with the cluster and the cluster name, the length of the minimum spanning tree, the length of total plant that the HAI builds within the main cluster, and the difference between the two (the shortage). This is only a sample, for illustrative purposes.

[redacted]	[redacted]	[redacted]	[redacted]
CPCRFLXA008	108,716	60,694	48,022
CLTNFLXA002	45,131	181	44,950
LBLLFLXA003	48,895	6,058	42,837
WCHLFLXA005	63,122	23,169	39,953
NPLSFLXC004	50,783	13,048	37,735
IMKLFLXA003	54,642	18,966	35,676
OKCBFLXA018	81,317	46,014	35,303
LKPCFLXA009	45,311	10,818	34,493
PTCTFLXA033	107,854	73,536	34,318

As you can see, the lengths that the HAI Model underbuilds are not insignificant. In the table below, I list the total in miles of this underbuilding, by density zone, for Sprint's serving territory. Recall, the shortage listed on each line below does not address outlier clusters, nor does it address feeder in any way. The shortages listed are found within main clusters.

[redacted]	[redacted]	[redacted]	[redacted]
Sprint-United	**0 to 5**	637 miles	460 miles
Sprint-Centel	**0 to 5**	333 miles	223 miles
Sprint-United	**5 to 20**	434 miles	288 miles
Sprint-Centel	**5 to 20**	669 miles	381 miles
Sprint-United	**20 to 100**	244 miles	138 miles
Sprint-Centel	**20 to 100**	91 miles	39 miles

Q. Have results similar to these been found in other states?

A. Yes. In every state for which Sprint has seen the actual cluster data and been able to perform such an analysis, the result is always the same: In the low density areas, this underbuilding is systematic, significant, and occurs in the overwhelming majority of main clusters.

Q. How have the HAI Sponsors responded to these statements when presented to them?

A. In the aforementioned Texas proceeding, the response of the HAI proponents was twofold. First it was suggested that if this is indeed a problem or shortcoming for the HAI Model, the BCPM would exhibit the same shortcoming to a much greater degree. Mercer Supplemental Testimony, June 5, 1998, states "Sprint's claim of a flaw is misleading, greatly overstated, and is of equal or more applicability to the BCPM as well."

Second, the HAI proponents claimed that the HAI built substantially more backbone and branch cable inside their clusters than the BCPM built inside its grids, a statement which was intended to support the first statement above. (Mercer/Klick, June 30, 1998.)

Q. Have the BCPM Sponsors conducted a MST analysis on their own model in Florida?

A. Yes we have. It is not possible to replicate the exact MST analysis that was done on the HAI Model because the BCPM in its standard format does not place points, but places counts of customers within microgrids. However, if assumptions are made regarding how these counts are placed in a microgrid, it is possible to conduct a type of MST analysis that measures the dispersion of original customer locations and how that compares with the cable built by the BCPM. A discussion of the BCPM MST approach is attached as Exhibit BKS-10.

For the HAI Model, our analysis was done at the main cluster level. The equivalent level in the BCPM is the ultimate grid level, and this is the level that was used for our MST analysis. The table below shows the results for the same density zones as shown above for the HAI Model, for all of Sprint's territory in Florida. (Due to time constraints I was unable to separate grids by company.)

[redacted]

0 to 5	1,164	335	28.8%	171	14.7%

5 to 20	787	89	11.3%	25	3.2%
20 to 100	721	4	< 1%	2	<1%

As the table shows, there is evidence that sometimes the BCPM underbuilds in rural Florida. However, the frequency of this occurring is much smaller than with the HAI Model. Using 87% of the Minimum Spanning Tree as the measure of what is "sufficient" cable, recall that the HAI Model underbuilt well over 85% of main clusters. By comparison, the BCPM underbuilds less than 15% of grids. The HAI Sponsors' claim, that the BCPM exhibits the same problem to an equal or greater degree, is without foundation.

In addition, it can be worthwhile to compare actual plant built by each model within the basic unit of analysis, either the main cluster (for HAI) or the ultimate grid (for BCPM). Unfortunately, the two units do not directly equate to each other, so any meaningful comparisons must be made at the wire center level, and even then the comparison is imperfect. First, because our analysis focuses only on main clusters, it would be incorrect to compare a HAI total with a BCPM total for the same wire center. Second, examining data at the wire center level misses important detail because it allows high-density areas within the wire center to offset low-density areas. The solution is to look at wire centers that are low-density overall.

In the table below we provide the following information:

What the HAI Model builds within main clusters for an entire wire center;
What the total MST distance is for the main clusters in that same wire center;
The degree, if any, to which the HAI Model fell short of "sufficient" cable;
What the BCPM builds within ultimate grids for an entire wire center;

What the total MST distance is for the ultimate grids in that same wire center;

The degree, if any, to which the BCPM fell short of "sufficient" cable.

As stated earlier, our concern is with the lowest density areas of Florida, since these are clearly of highest concern for universal service purposes. The table lists the wire centers, in Sprint's Florida serving territory, where the overall density was less than 20 lines per square mile.

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GNVL	837,911	1,241,375	403,464	1,574,751	1,321,860	0
GLDL	547,884	695,539	147,655	986,196	791,228	0
PNLN	758,103	995,501	237,398	1,185,130	1,036,367	0
LEE	641,367	966,026	324,659	1,304,735	1,173,921	0
KNVL	844,510	310,829	0	863,493	605,045	0
ZLSP	1,181,784	995,511	0	1,312,056	1,103,090	0
SPCP	496,392	694,267	197,875	969,965	781,920	0
CHLK	1,008,642	1,313,833	305,191	1,673,651	1,436,335	0
RYHL	658,109	896,039	237,930	1,167,481	956,386	0
GNWD	875,148	976,640	101,492	1,562,988	1,352,350	0
EVRG	744,918	327,307	0	505,130	386,073	0
MALN	694,647	806,258	111,611	1,184,506	1,056,533	0
BAKR	1,447,839	1,547,207	99,368	2,595,212	2,059,406	0
FRPT	1,049,030	1,268,181	219,151	1,984,645	1,389,764	0
MNTI	2,507,994	2,941,833	433,839	4,395,127	3,469,573	0
CTDL	590,714	580,683	0	948,482	721,563	0
WSTV	68,129	85,375	17,246	94,145	76,766	0
GDRG	800,128	759,808	0	1,319,982	1,044,484	0
STMK	170,084	241,346	71,262	430,952	333,115	0

As the table shows, in every case where a reasonably direct comparison can be made, the

BCPM builds sufficient plant at the wire center level, whereas the HAI falls short in the majority of cases.

Also, it is important to note that in many of the wire centers shown above, the MST distances are roughly similar between the two models. The fact that MST lengths would be similar, but the HAI builds less than the MST while the BCPM builds more, lends support for the following: *A key difference between the two models is not merely how each model's preprocessing initially allocates customer locations. Rather it is in how closely the model comes to using those locations when it builds its network.*

Q. Is the plant listed in this table all categorized as distribution plant, or backbone and branch cable?

A. No, not for either model. Connecting cable is included in the table above because, in both models, connecting cable is built inside the basic unit (the grid or the cluster) to connect customers in one section of the grid/cluster with customers in another section. In the BCPM it is used more often than in the HAI.

There has been a great deal of confusion as to what types of plant or cable should be included when calculating "what either model builds". For the HAI Model, in the majority of cases the basic unit of analysis, the cluster, represents one serving area and one distribution area. In the BCPM, most ultimate grids represent one carrier serving area that is separated into (up to) four distribution areas. Sometimes the two models have different terms for the cable that is used at various points in the network. Because of this, it is best to consider all plant built within the basic unit (grid/cluster) since all of it may be used for the purposes of connecting customers to each other and to the network.

In most cases for the HAI, a customer on the west side of any cluster is connected to a customer on the east side of that cluster using *backbone or branch cable*. The same two customers would be connected in the BCPM using *connecting cable*. Both are copper.

On the other hand, connecting cable in the BCPM often is found on the customer side of the DLC. Connecting cable in the HAI Model is usually found on the office side of the DLC. In the HAI Model, connecting cable is fiber. In the BCPM it is not.

Because of this potential for confusion, the comparisons above used everything that exists solely within the cluster or grid: connecting cable plus distribution (backbone and branch) cable. Drop was also included. The results of the table demonstrate the following: A proper comparison of the amount of plant built by the two models, eliminating any confusion over nomenclature or terminology, demonstrates that the shortages discovered in the HAI are significant and systematic in the rural areas of Florida, while the BCPM does not suffer from the same shortcoming.

Q. Please summarize your rebuttal testimony.

A. In his direct testimony Mr. Don Wood states that a model must do two things: It "must accurately determine customer locations" and it must "connect those customers with the serving central office using network facilities that are efficient…" (Wood Direct page 4).

In my rebuttal testimony I have clarified the first point. Contrary to Mr. Wood's statement, a model must not only "determine" customer location but use the location as well. We have seen that the HAI Model does not do this.

Second, according to Mr. Wood the model must "connect" the customers to the network. The analysis presented here provides evidence that in the rural areas of Florida the HAI Model fails this test as well.

I have shown that the HAI Model consistently and significantly underestimates and underbuilds the amount of cable needed to do exactly what Mr. Wood states it must, "connect" customers to the network. I have presented evidence that the FCC conducted its own analysis that supports the findings shown here. In summary, the HAI Model is not the most accurate and reliable costing methodology available to the Commission but a model that is fundamentally and systematically flawed.

Q. Does this conclude your rebuttal testimony?

A. Yes.

EXHIBIT BKS-1A
FOOTNOTES

[1] A detailed explanation of *how* this underbuilding occurs is contained in the Sprint ex parte April 17, 1998.

[2] Geocoded locations also figure into the HAI preprocessing in that the *convex hull* and *minimum bounding rectangle* of any group of points is used to create an artificial polygon which enters the HAI Model as a cluster. However, it is incorrect to say that this artificial polygon retains any of the pertinent information contained in or reflected by the original geocoded data.

[3] Sprint's analysis has always focused on *main* clusters only. Sprint has never commented or conducted analysis on the HAI's *outlier* clusters in any jurisdiction, with the exception of pointing out that outlier clusters in the HAI are served using T1 repeaters over copper and T1 does not constitute a forward-looking methodology.

[4] Actually, connecting cable in the HAI is used to connect two or more DLCs and so technically it does not fall on the customer side of the DLC. However, because it is used to span distances between customers within main clusters, we feel it is appropriate to include it when calculating "what the HAI Model builds".

Exhibit BKS-2

Sprint

Pete Sywenki
Director, Federal Regulatory Relations

Law & External Affairs
1850 M Street, NW, Suite 1100
Washington, DC 20036
Voice 202 828 [illegible]
Fax 202 296 3469
pete.n.sywenki@mail.sprint.com

EX PARTE

April 17, 1998

STAMP & RETURN

RECEIVED
APR 17 '98
FEDERAL COMMUNICATIONS COMMISSION
OFFICE OF SECRETARY

Ms. Magalie Roman Salas
Secretary - Federal Communications Commission
1919 M Street, N.W. Room 222
Washington, D.C. 20554

RE: CC Docket Nos. 96-45 and 97-160

Dear Ms. Salas,

On April 16, 1998, Jim Sichter, Brian Staihr, and Pete Sywenki of Sprint met with Brian Clopton, Natalie Wales, Brad Wimmer, Don Stockdale, Chuck Keller, Craig Brown, and Richard Smith of the Universal Service Division of the Common Carrier with regard to the above referenced dockets. In this meeting, we discussed the status of the cost proxy model platforms currently under the FCC's consideration for use in determining universal service funding for high cost areas. The attached information was discussed in the meeting. These attached materials illustrate the methodology by which customer locations are converted into serving areas for use in the HAI model and point out the way in which this approach significantly understates required distribution facilities.

Included in these materials are estimated distances (lengths) between customer location points within specific clusters. The calculation of these distances was developed by Sprint staff during an on-site review of PNR geo-coded data at PNR Associates (the vendor used by HAI for customer location points and clustering). This review was arranged in response to Sprint's requests during recent Nevada PUC costing proceedings. The information provided during our April 16 meeting did not include any actual customer locations or any other information proprietary to PNR.

The original and three copies of this notice are being submitted to the Secretary of the FCC in accordance with Section 1.1206(a)(1) of the Commission's rules. If there are any questions, please call.

Sincerely,

Peter N. Sywenki.

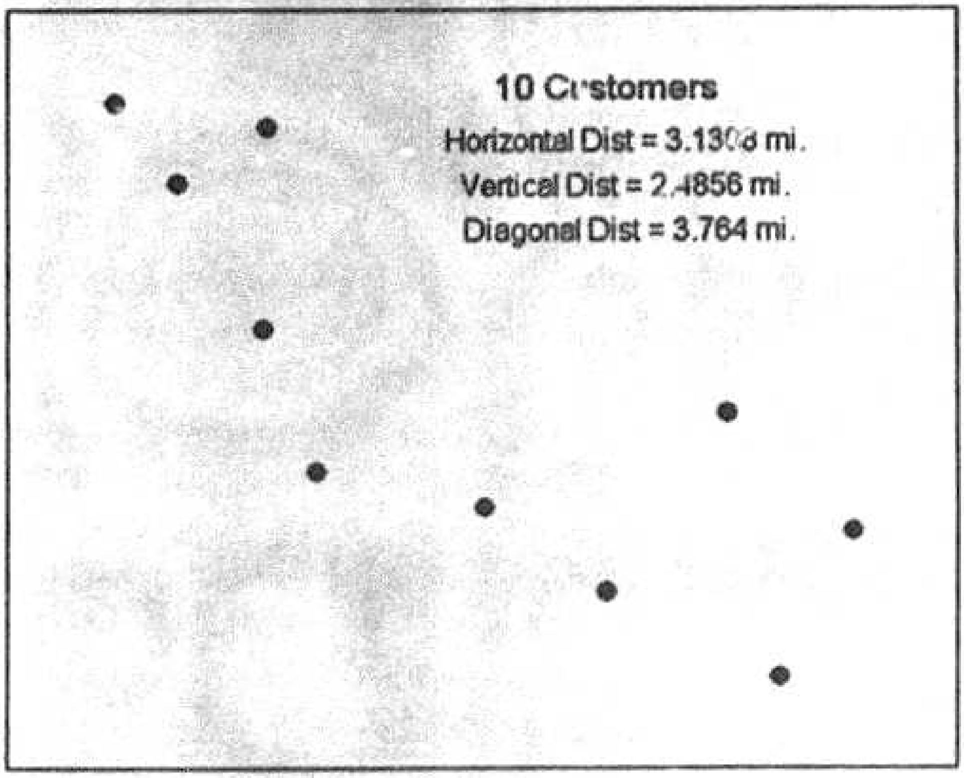
Pete Sywenki

Attachments

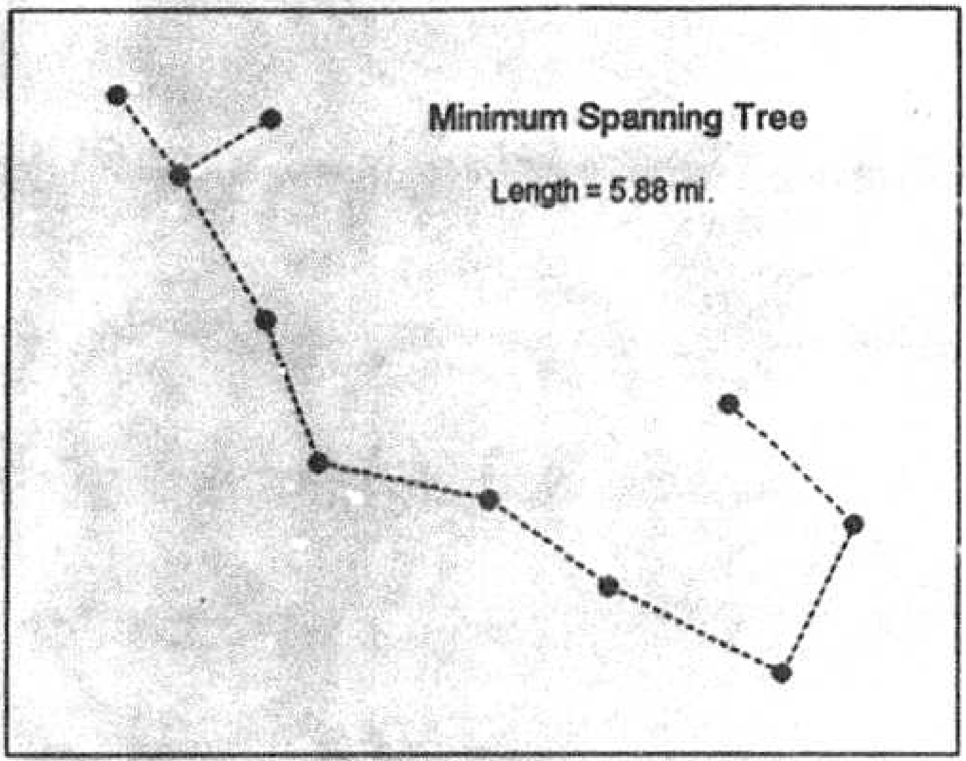
cc:
Brian Clopton
Natalie Wales
Richard Smith
Chuck Keller
Don Stockdale
Brad Wimmer
Craig Brown

Hatfield's Polygons Converted to Rectangles

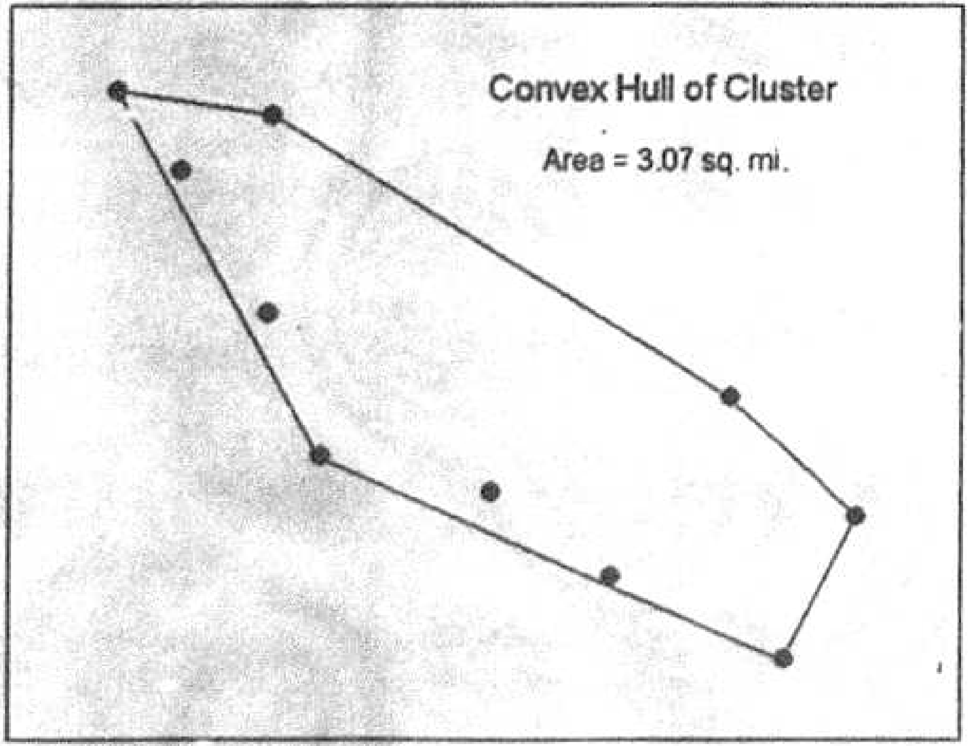
The Hatfield 5.0a Model groups a set of "actual" customer points into a *cluster*, according to a set of aggregation rules.



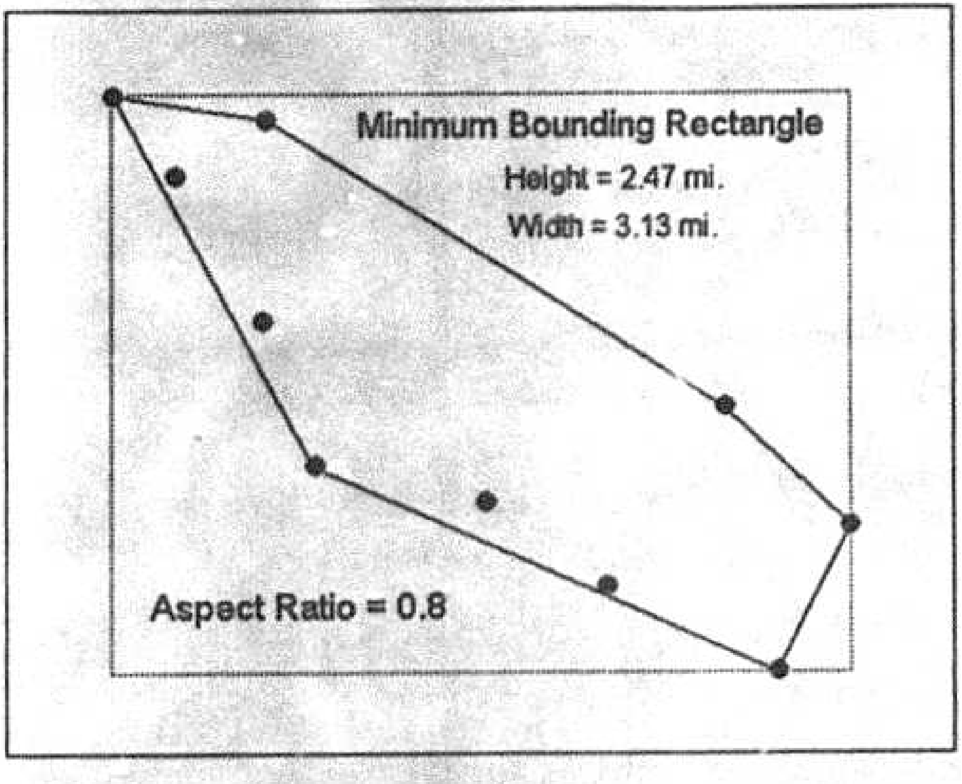
We have determined that the *minimum spanning tree* for these points – the mathematically shortest connection possible for these points – is 5.88 miles.



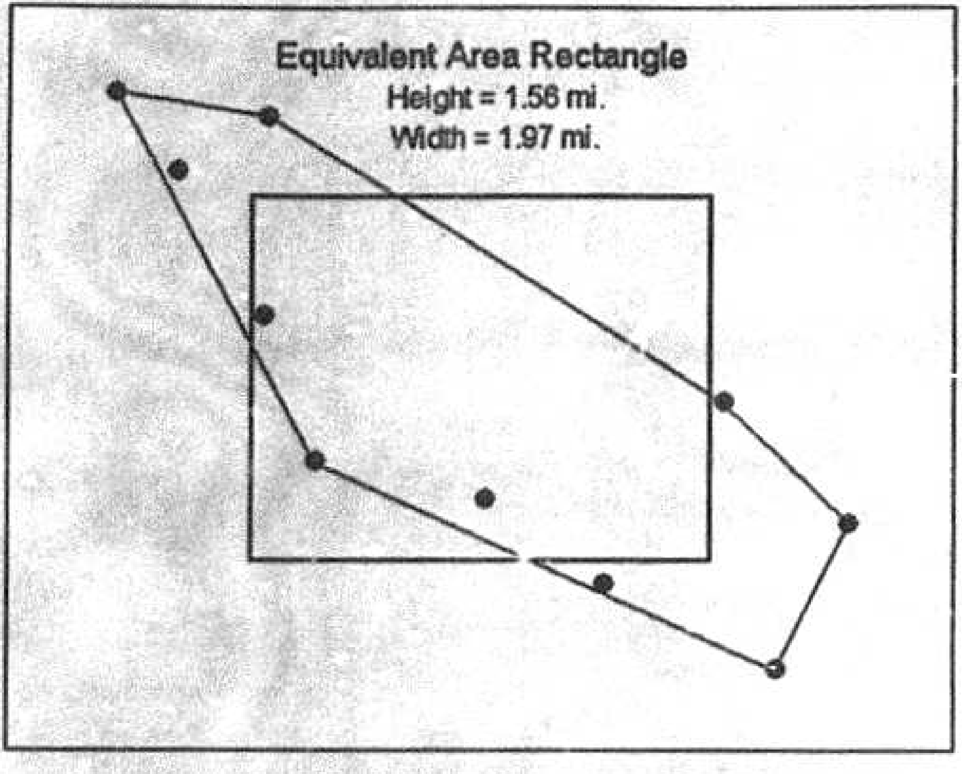
When Hatfield has determined the set of points that constitute a cluster, it logically draws a *convex hull* around those points, and determines its area.



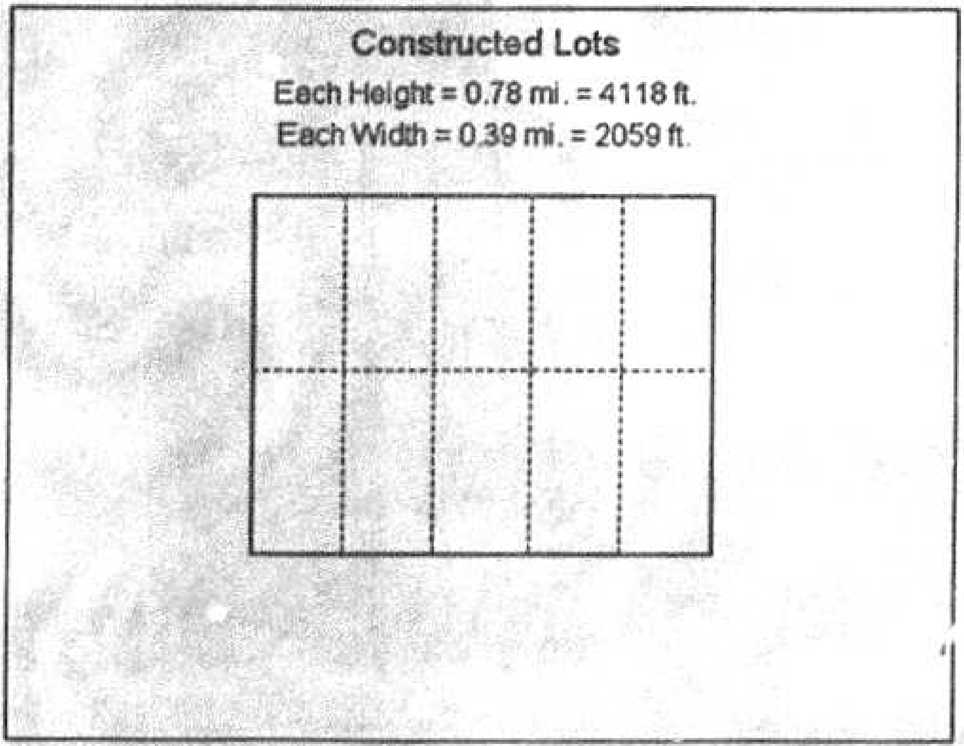
Hatfield then logically constructs a ***minimum bounding rectangle*** – oriented north-south-east-west – that exactly bounds the cluster's points. Hatfield then determines the ***aspect ratio*** of that rectangle (that is, the ratio of the rectangle's height to its width) … in this case, 0.8.



Hatfield then constructs a *rectangle* with the above aspect ratio; the *size* of that rectangle is determined, of course, by its *area* … and that area is set to be the *area of the convex hull* … in this case, 3.07 square miles.



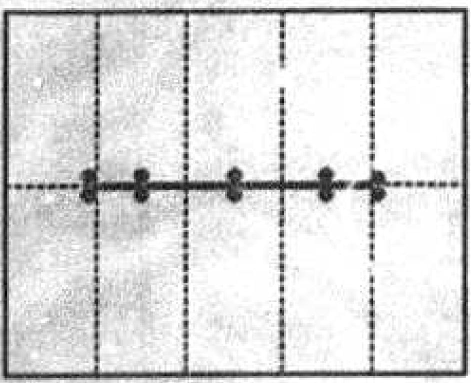
Hatfield then constructs *lots* within this constructed rectangle. Each lot is twice as high as it is wide.



A *branch cable* is then constructed, and 150 ft. drops connect to the customers.

Cabling to Serve Customers

Branch Cable Length = 6177 ft.
10 Drops, each at 150 ft.



Total Cable Length = 7677 ft. = 1.45 mi.

Less than 1/4 of the Minimum Spanning Tree length!

But note how closely the customers are squeezed toward the branch cable. The arrangement is unrealistic, both from the standpoint of cable length *and* from the standpoint of area served.

Customer Area Served

Height = 300 ft.

Width = 106 + 6177 + 106 ft. = 6389 ft.

Area Served = 1,916,700 sq. ft. = 0.0688 sq. mi.

But Actual Cluster Area = 3.07 sq. mi.

Area Modeled is 1/44 of Cluster Area

So, HOW BAD CAN THIS BE?

To what extent does the combined effect of:

1) converting the polygon into a rectangle (with identical area) and
2) building cable only to the point where the perimeter lots start
3) assuming all customers have drops 150 feet or less

cause the model to UNDERSTATE the amount of cable needed to transverse the ACTUAL distances between customers?

The following table shows a sample of several individual clusters (not wire centers) in Nevada (Nevada Bell territory).

The table gives an example of the amount of cable needed to reach all actual customer locations in the cluster. The locations do NOT include any outlier locations. The distance reported is only the distance between points that reside in the main clusters.

This length represents an approximation of the amount of distribution that the Hatfield Model (or any proxy model) should build in the course of laying out the network and determining the associated cost.

The table also shows the amount of actual distribution the Hatfield Model builds to each respective cluster (again, excluding outlier points).

Cluster Number	Absolute Minimum Distance Between Cluster Points (in feet)	Total Amount of Distribution Cable Built by Hatfield Model (in feet)
CHBTNV11.C003	23,500	7,900
IMLYNV12.C022	29,000	2,210
UPMDNVXF.C005	29,000	836
IMLYNV12.C015	38,000	2,089
DYTNNV11.C004	21,000	1,494
EMPRNV11.C004	21,500	5,093
EMPRNV11.C003	24,500	0

Sprint

Pete Sywenki
Director, Federal Regulatory Relations

Law & External Affairs
1850 M Street, NW, Suite 1110
Washington, DC 20036
Voice 202 828 7452
Fax 202 296 5469
pete.n.sywenki@mail.sprint.com

EX PARTE

April 23, 1998

Ms. Magalie Roman Salas
Secretary - Federal Communications Commission
1919 M Street, N.W. Room 222
Washington, D.C. 20554

RE: CC Docket Nos. 96-45 and 97-160

Dear Ms. Salas,

Today, I sent the attached information to Chuck Keller and Brad Wimmer of the Universal Service Division of the Common Carrier Bureau in regard to the above referenced dockets. This information supplements and clarifies issues shared with the staff in an April 16, 1998 meeting concerning the clustering and distribution methodology employed by the HAI model.

The original and three copies of this notice are being submitted to the Secretary of the FCC in accordance with Section 1.1206(b)(1) of the Commission's rules. If there are any questions, please call.

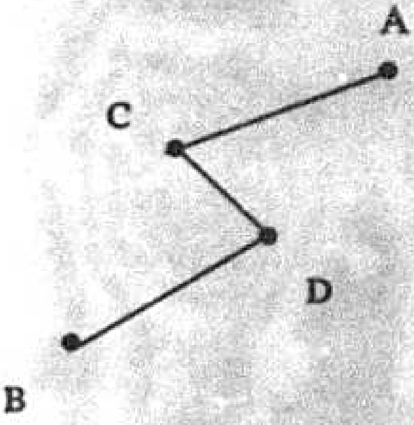
Sincerely,
Pete N. Sywenki
Pete Sywenki

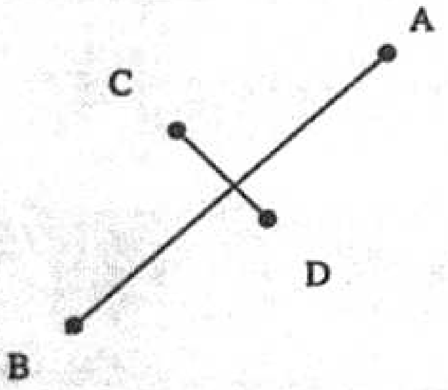
Attachments

cc: Chuck Keller
Brad Wimmer

Hatfield Clusters and a Note Regarding the Minimum Spanning Tree:

In a few unique cases, it is possible that points in a telephone network could be connected with an amount of cable that was slightly less than the length of the minimum spanning tree. An example of this is shown below.





The cable connecting the 4 points in the left figure represents the minimum spanning tree, and is slightly longer than the cable length in the figure on the right. (In actual measures, the cable in the right figure is 92.3% of the length of the cable in the left figure.)

In the few cases where this might occur, we have found the difference in length to be consistently less than 10%. In other words, where a minimum spanning tree for a given cluster might equal 10,000 feet, it is possible that the points in that cluster could be served with only 9,000 feet of cable (depending on how those points were configured).

To eliminate confusion, the table on the following page lists several additional Nevada clusters and the length of each cluster's longest diagonal. (In the examples above, this would be the distance between points A and B.) The table also lists the amount of distribution cable built by the Hatfield Model to serve these clusters.

The cluster's longest diagonal serves as an absolute lower bound in terms of the amount of cable required (since the cluster's configuration is based on the actual locations of the points in that cluster.) The only case in which the diagonal length would represent the actual required length of cable is when all points in a cluster were located in a straight line.

In the vast majority of cases, the minimum required cable would be significantly MORE than the longest diagonal of the cluster. But using the diagonal length provides us with an ultra-conservative measure of the required plant per cluster.

Hatfield 5.0a Cluster Diagonal / Distribution Comparison

Cluster Name		Approximate Length of Polygon Diagonal (Prior to Conversion to Rectangle. In feet.)	Total Distribution Distance within Main Cluster built by Hatfield Model 5.0a (Distribution Module Cell BU minus Cell CQ (Cell CQ represents outlier road distance, in most cases zero))
ALAMNVXF	C001	31,000	40,274
"	C002	27,000	3,463
"	C003	35,000	958
"	C004	25,000	13,257
"	C005	35,000	18,517
"	C006	22,000	5,296
"	C007	28,000	14,523
"	C008	25,000	22,296
"	C009	26,000	10,289
"	C010	25,000	15,922
"	C011	23,000	6,653
"	C012	12,000	2,141
"	C013	29,000	22,008
"	C014	28,000	10,781
"	C015	18,000	12,344
"	C016	25,000	7,719
"	C017	27,000	7,471
"	C018	28,000	8,697
AUSTNV11	C001	22,000	9,276
"	C002	21,000	1,305
"	C003	27,000	354
"	C004	34,000	13,911
"	C005	28,000	4,208
"	C006	22,000	1,481
"	C007	19,000	1,657
"	C008	19,000	2,755
"	C009	24,000	68,331

Note that out of 27 clusters in 2 wire centers shown above, only two (2) clusters built more distribution than the cluster diagonal.

It is important to remember that this is no way implies that sufficient distribution was built in those two clusters, only that sufficient distribution was built to cross the cluster's diagonal axis. For example, in the cluster ALAMNVXFC001, the Hatfield Model builds 46 lines and in the cluster AUSTNV11C009, the Hatfield Model builds 215 lines. The large number of lines in each suggests a need for substantially more distribution than just the amount required to traverse the cluster's diagonal.

Sprint

Pete Sywenki
Director, Federal Regulatory Relations

Law & External Affairs
1850 M Street, NW, Suite 1100
Washington, DC 20036
Voice 202 828 7452
Fax 202 296 3469
pete.n.sywenki@mail.sprint.com

<u>EX PARTE</u>

April 30, 1998

Ms. Magalie Roman Salas
Secretary - Federal Communications Commission
1919 M Street, N.W. Room 222
Washington, D.C. 20554

RE: CC Docket Nos. 96-45 and 97-160

Dear Ms. Salas,

Today, I provided the attached materials in regard to the above referenced dockets to Brad Wimmer and Chuck Keller of the Universal Service Branch of the Common Carrier Bureau. The attached materials are related to a potential correction to the HAI model's current clustering methodology. This was suggested as part of an ex parte submitted by the HAI model sponsors on April 23, 1998. As shown in the attachment, the suggested HAI model "fix" would still result in a significant shortfall in the amount of distribution cable that would be needed to connect all of the customer locations in a cluster.

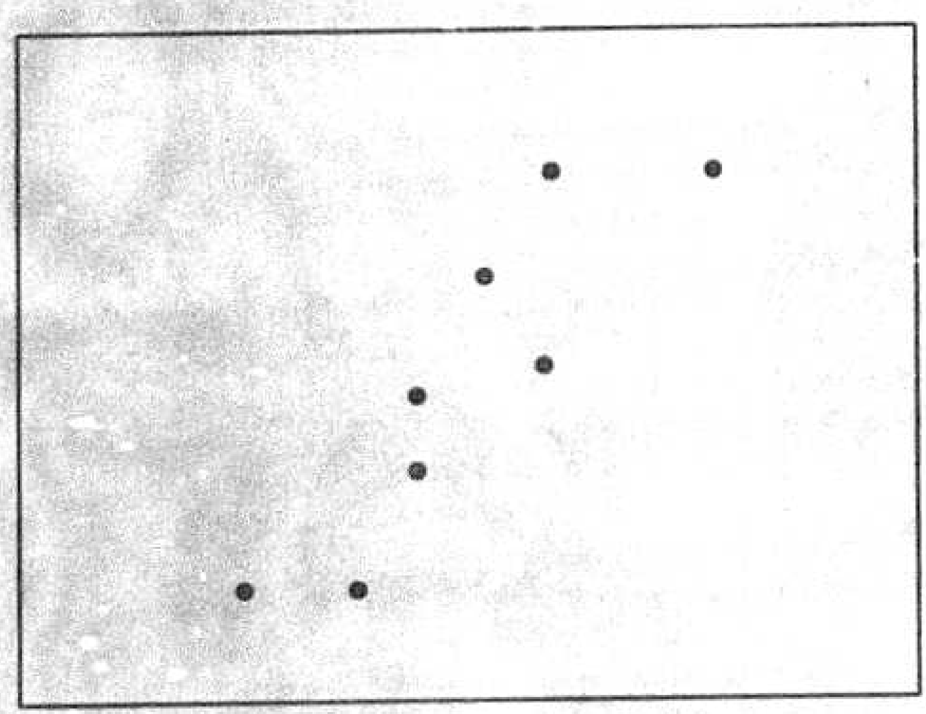
The original and three copies of this notice are being submitted to the Secretary of the FCC in accordance with Section 1.1206(b)(1) of the Commission's rules. If there are any questions, please call.

Sincerely,

Peter N. Sywenki

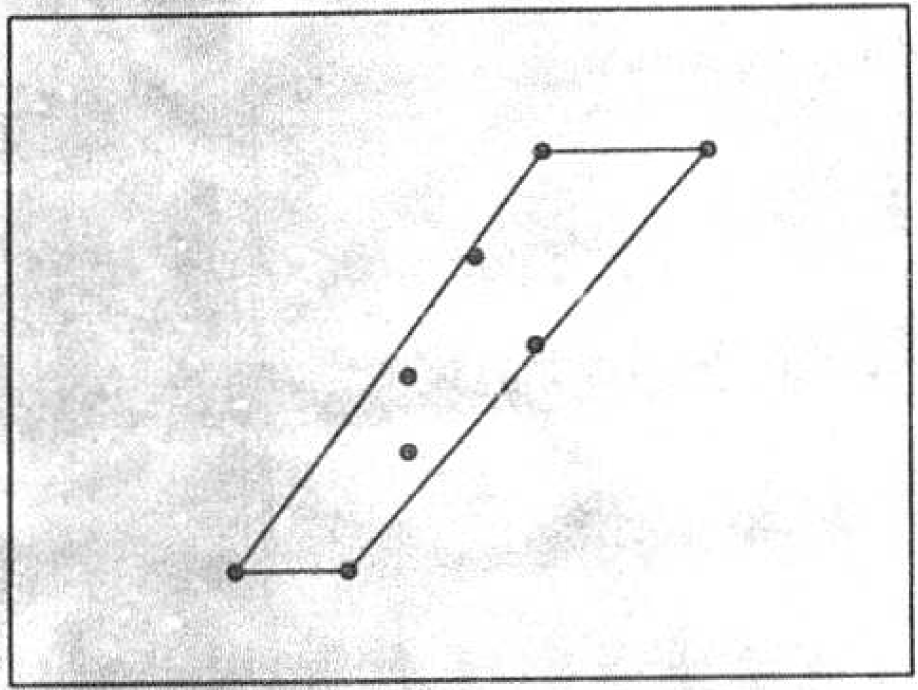
Pete Sywenki

cc: Chuck Keller
Brad Wimmer

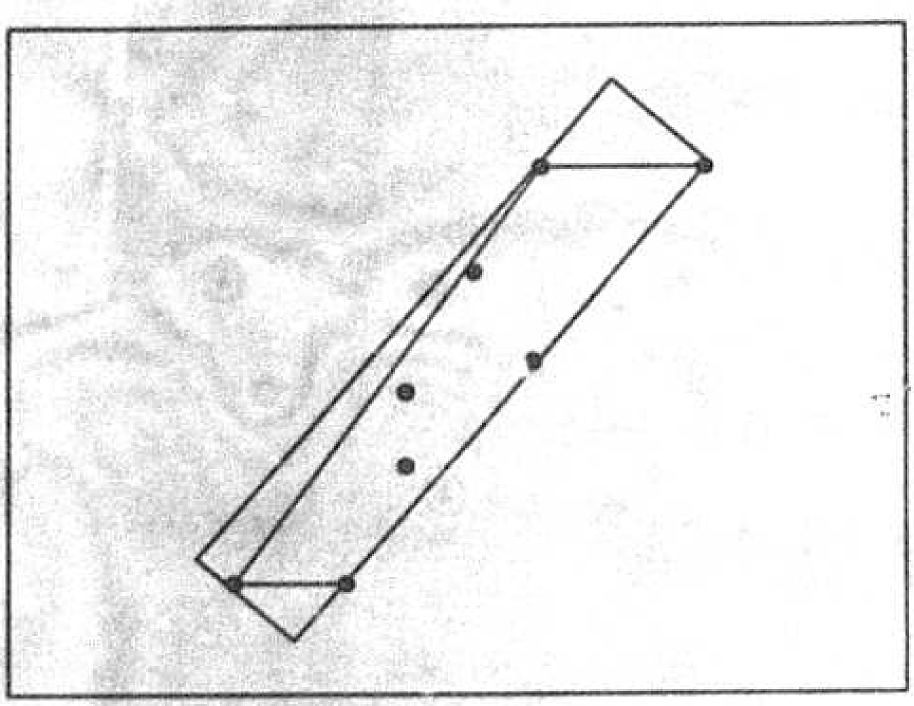


Original Points, representing customer locations: Either actual geocoded points or surrogate locations, or a combination of both.

These points will be grouped into one cluster according to PNR's clustering algorithm.



The convex hull containing the points on the previous page. The area of this polygon is converted to the rectangle.

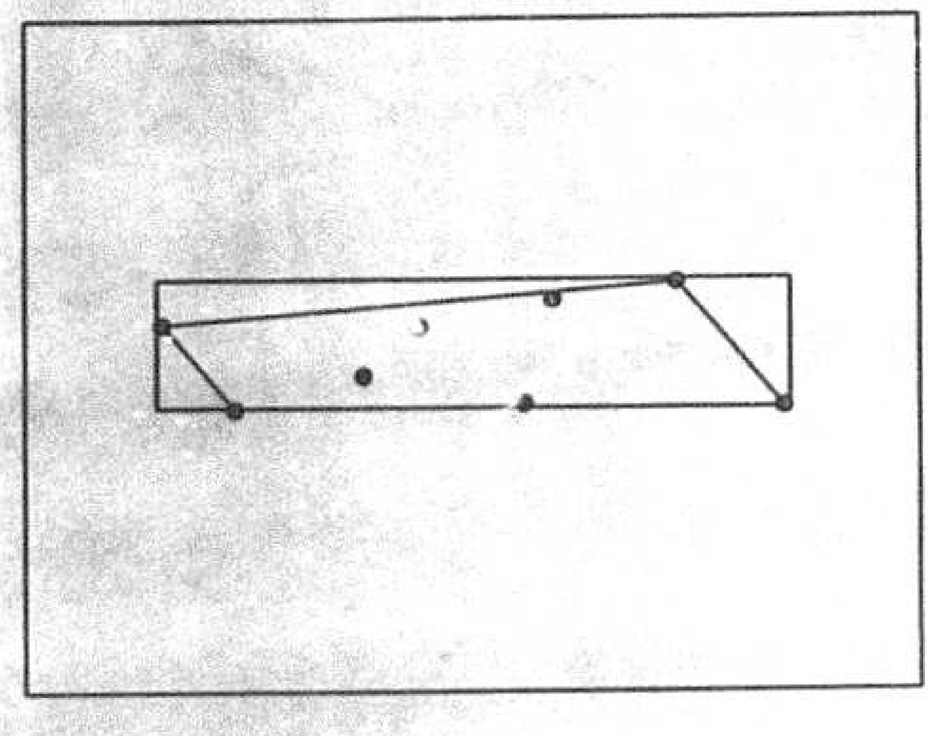


HAI Sponsor's Suggested "Fix"

Take the rectangle that actually surrounds the convex hull and maintain THAT aspect ratio (as follows). This is the "rotation" mentioned in the HAI Sponsor's ex parte of April 23, 1998.

This approach differs from the current method, which (for this particular set of points) would produce a more square-like minimum bounding rectangle.

For ease of explanation, we tilt the rectangle level in the following pictures.

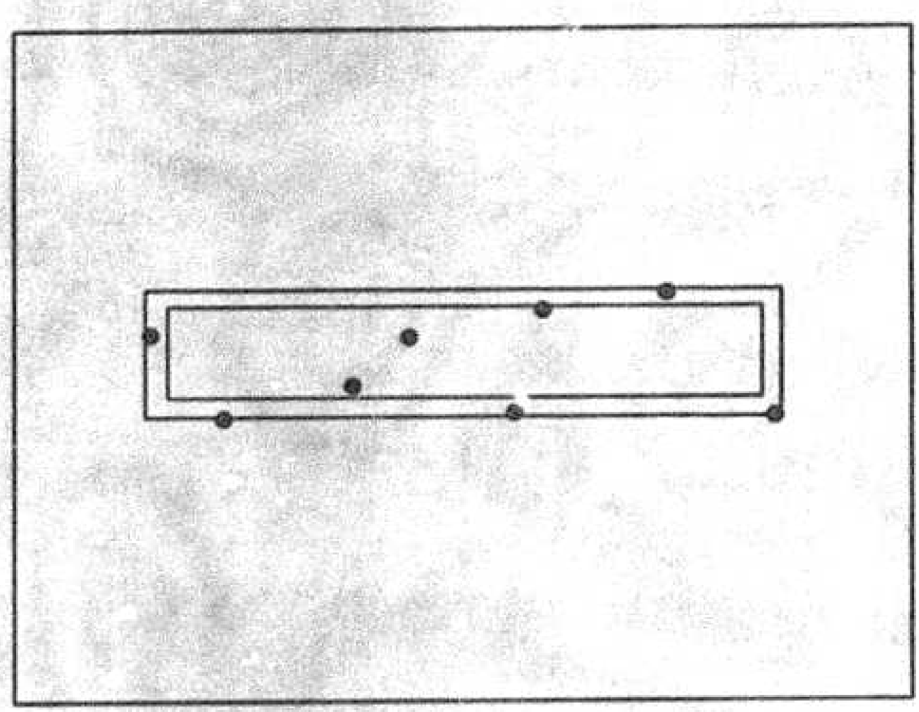


From this point there are two options:

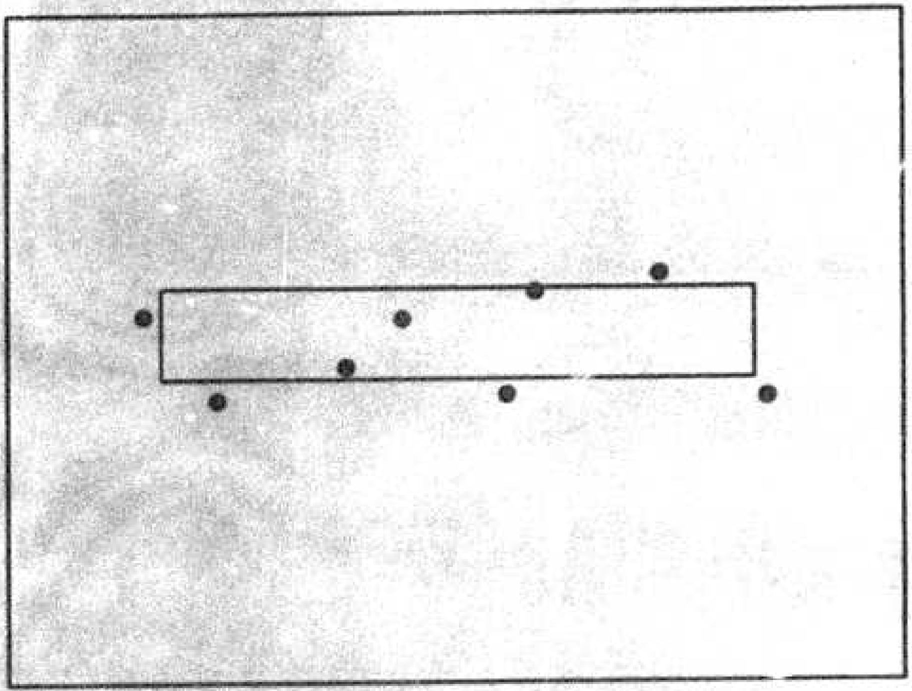
1) Maintain the area of the polygon using the aspect ratio of this surrounding rectangle. Build lots on that.

2) Maintain the area of this surrounding rectangle, build lots on that.

We produce option #1 first.

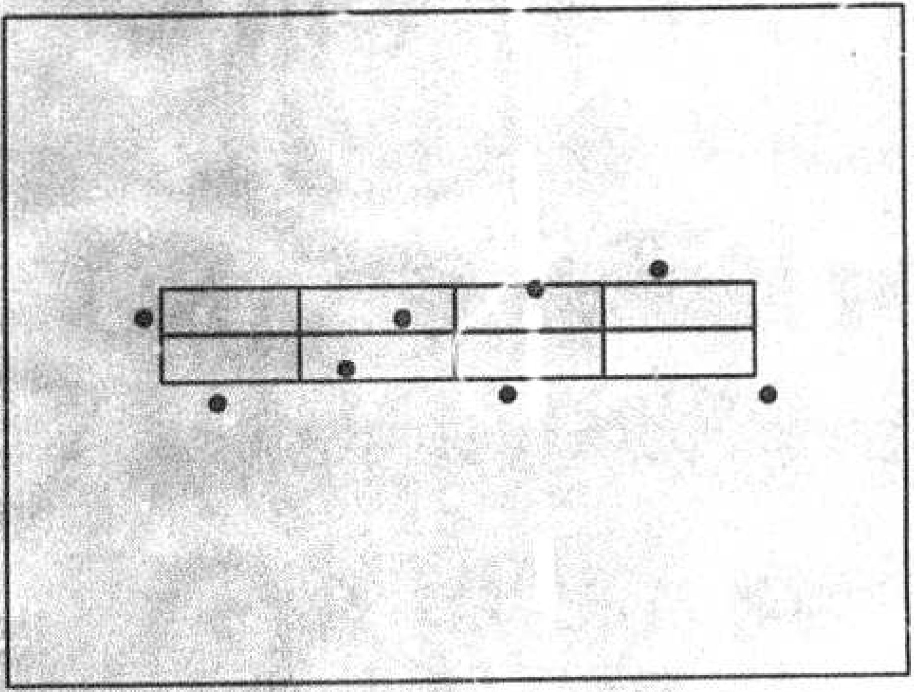


Smaller rectangle represents the area of the original polygon, converted to aspect ratio of the surrounding rectangle.



Polygon-converted-to-new-rectangle and original points.

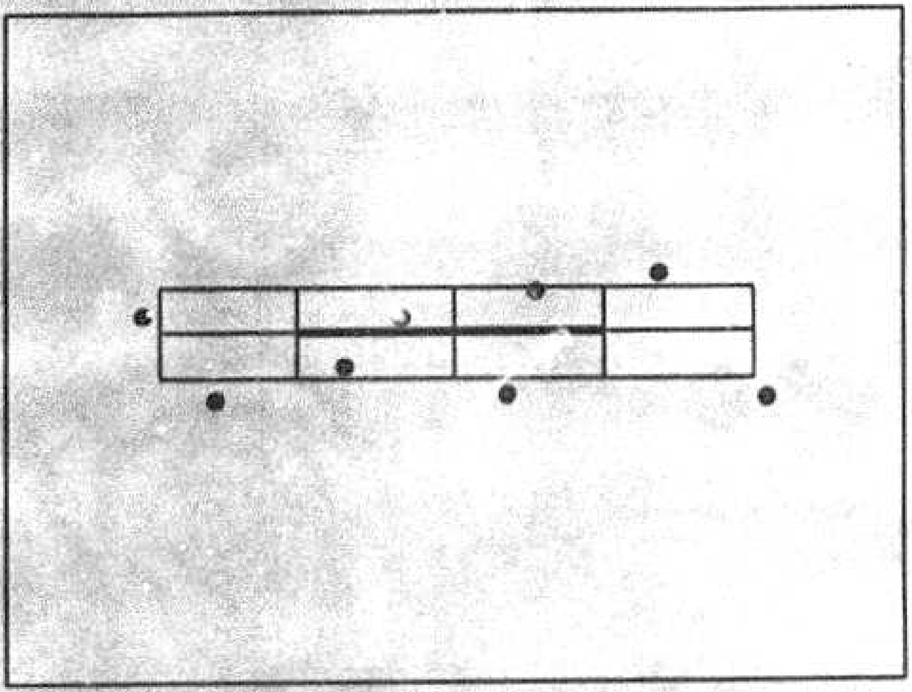
Already it is clear that the distance that must be covered between points is not encompassed in this area measure.



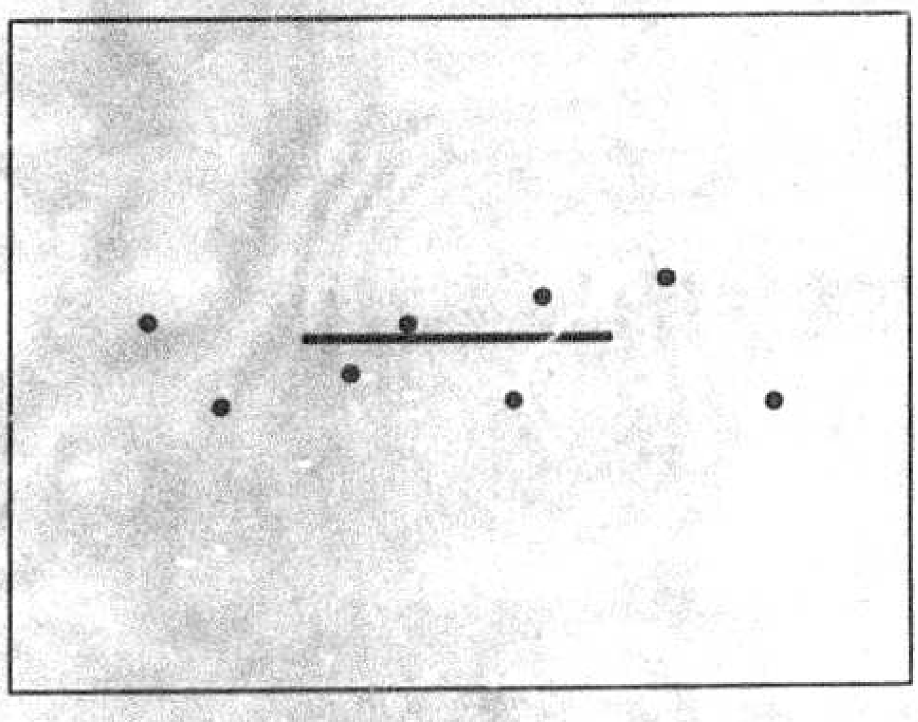
Converted area turned into lots, 8 lots for 8 customers.

Note: Lots are no longer exactly 2 X 1 (standard HAI 5.0a assumption). In maintaining the aspect ratio this no longer becomes possible.

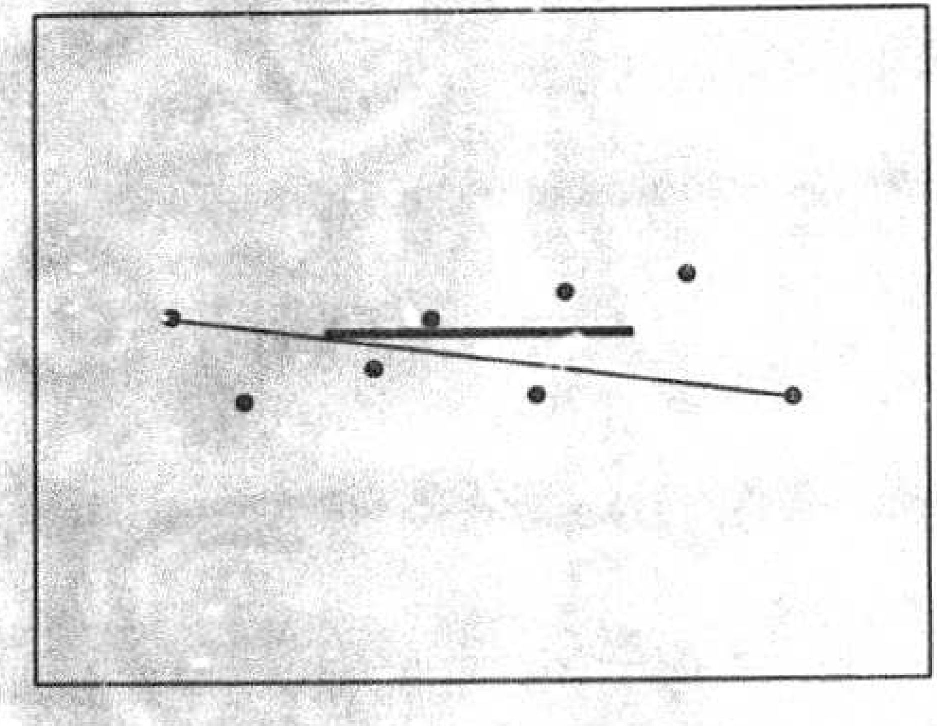
Feeder/distribution interface device will be placed at center of all 8 lots.



Red (or heavy) line represents total HAI distribution built to this cluster, even when rectangle is tilted and aspect ratio maintained.



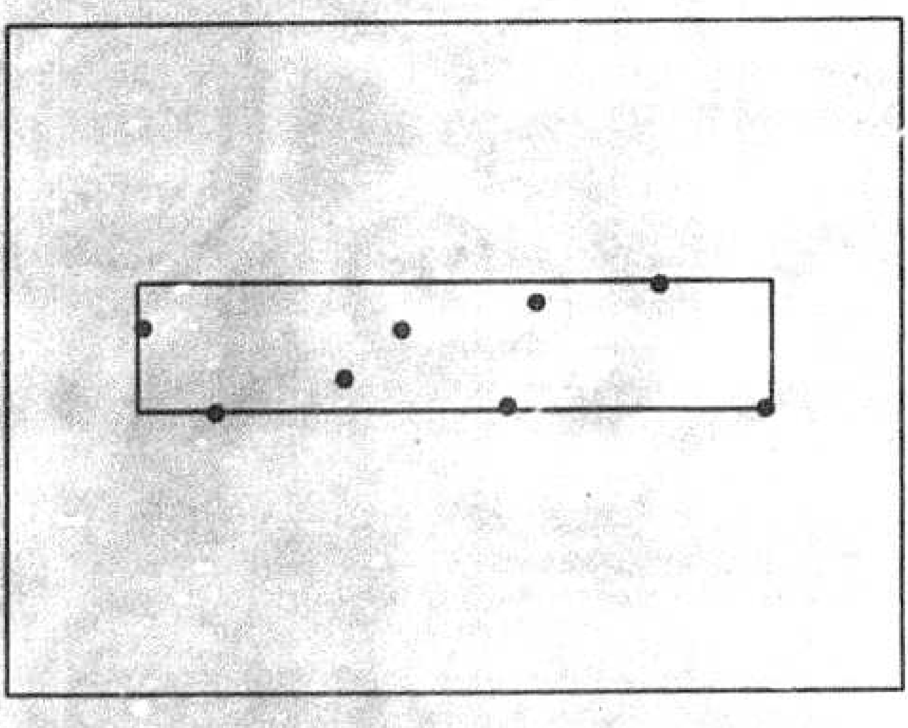
As picture shows, distribution falls dramatically short of distance required to connect customers where they are found.



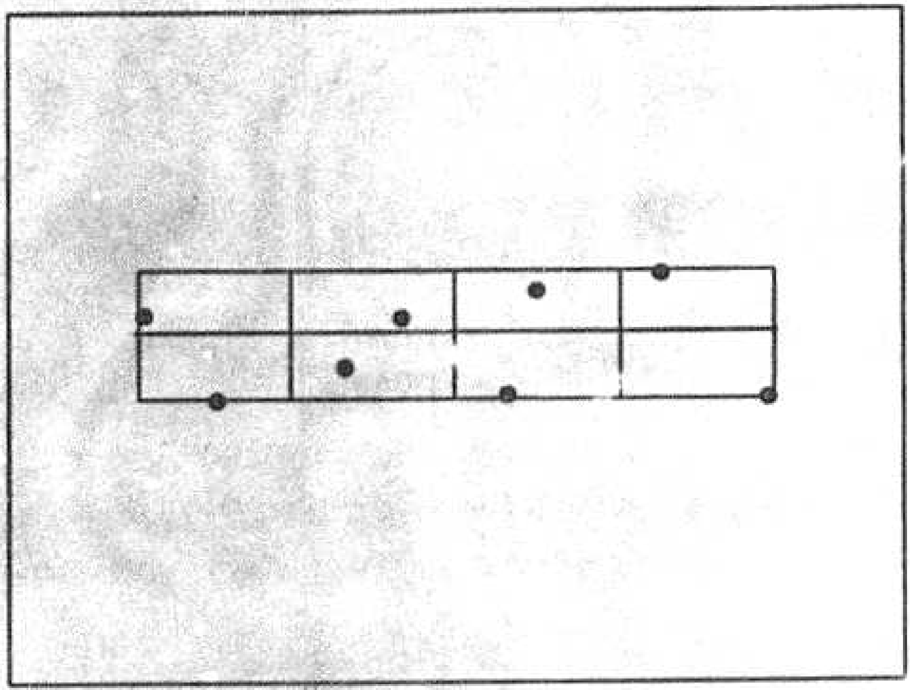
As picture shows, HAI distribution is dramatically short of even connecting the two customers that span the distance of the cluster.

This is the "diagonal" measure mentioned in Sprint's former ex parte.

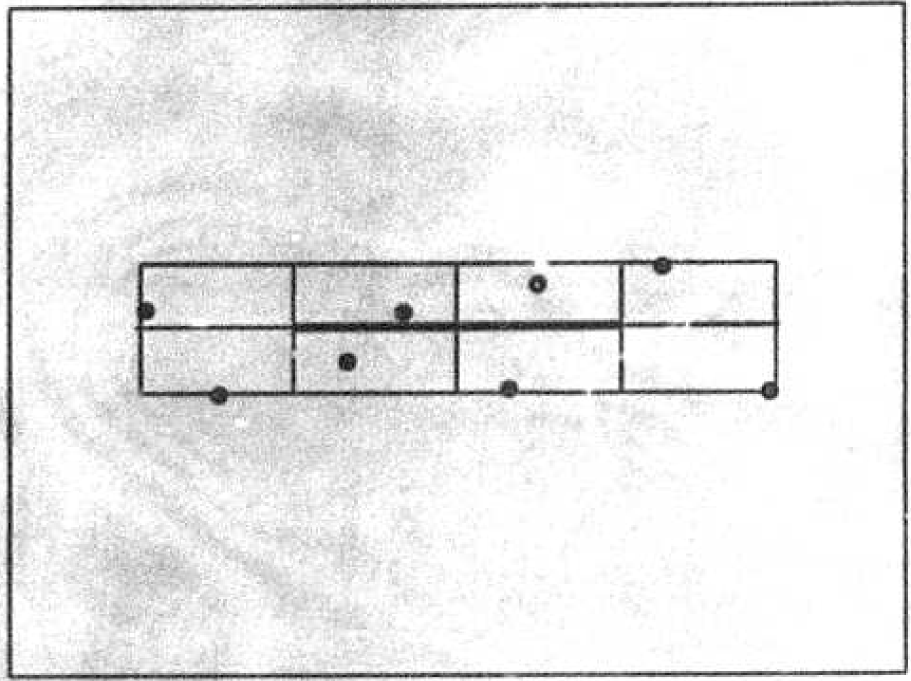
Now, option #2, maintaining the area of the tilted, bounding rectangle. (In most cases this will be larger than the area of the original polygon.)



Rectangle represents area of tilted minimum bounding rectangle, not area of convex hull of polygon.

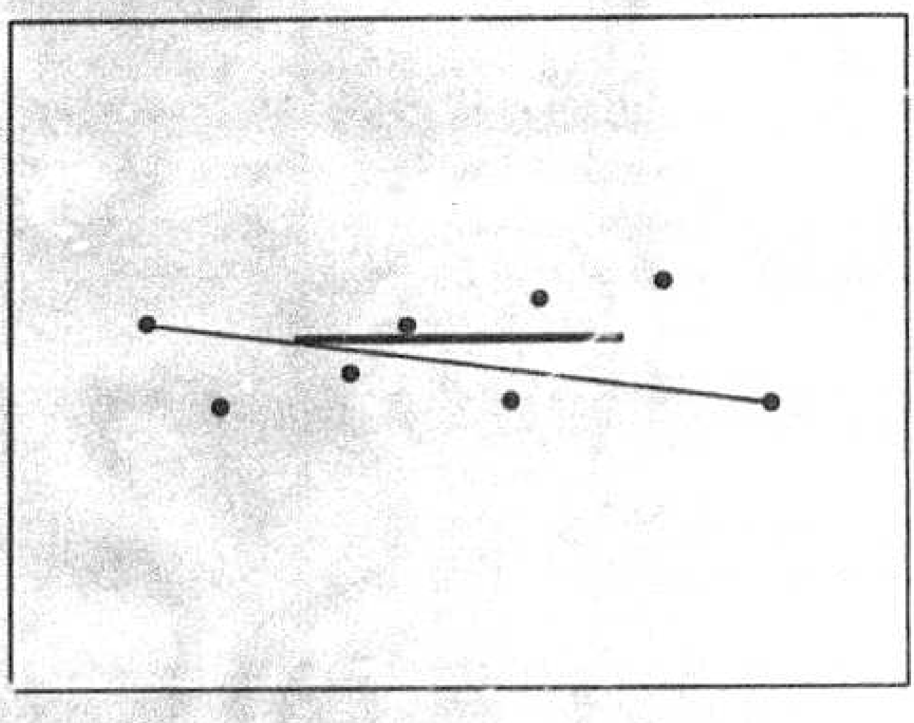


Lots placed over total rectangle area as before.



HAI Distribution cable built as before.

Key difference: Slightly larger lot size causes cable to be slightly longer than before.



Same effect: Cable falls far short of even the amount required to connect to farthest customers.

Important Notes:

1) There is NO offsetting effect.

2) As long as customer dispersion is condensed in any way, (such as converting polygons to rectangles or building cable only to the inside boundaries of perimeter lots), underestimation of required cable will occur.

3) "Distance between points", and not "area encompassed by points" must determine amount of cable built.

Exhibit BKS-3

EX PARTE OR LATE FILED

Federal Communications Commission
Washington, D.C. 20554

DOCKET FILE COPY DUPLICATE

May 13, 1998

RECEIVED

MAY 14 1998

FEDERAL COMMUNICATIONS COMMISSION
OFFICE OF THE SECRETARY

Richard N. Clarke
AT&T
Room 5462C2
295 North Maple Avenue
Basking Ridge, New Jersey 07920

Dear Mr. Clarke:

Attached please find an analysis prepared by staff members of the Common Carrier Bureau that deals with the HAI model's customer location algorithm. The analysis consists of a memorandum describing the analysis and a computer disk containing the results of the individual trials of the analysis. We would appreciate your comment on this analysis at your earliest convenience.

Sincerely,

Lisa S Gelb

Lisa S. Gelb
Chief, Accounting Policy Division
Common Carrier Bureau

Enclosures

cc: BCPM sponsors
Office of the Secretary, CC Docket Nos. 96-45/97-160

MEMORANDUM
UNITED STATES GOVERNMENT

Date: May 13, 1998

To: Magalie Roman Salas, Secretary

From: Jeffrey Prisbrey

Subject: A Test of Customer Dispersion in the HAI Customer Location Algorithm

---

On April 17, 1998 Sprint filed an ex parte containing an analysis of some customer location data inputs used in the HAI model for the state of Nevada. Sprint claimed that, for certain configurations of customer locations, the HAI model might understate the true cost of serving those customers because of the way in which the model converts the geocoded customer locations into a rectangular serving area. AT&T and MCI responded to the Sprint claims in an ex parte dated April 23, 1998, by claiming that "the situation identified by Sprint is rare, the total cost effect of correcting it is minor -- and likely to be at least partially offset by the effects of other HAI Model assumptions that have tended to overestimate costs."[1]

This memorandum documents the results of a test of the customer location algorithm used by the HAI model v5.0a that I have conducted as part of the ongoing staff evaluation of cost proxy models.[2] The test is designed to provide a measure of dispersion of customer locations both before and after an algorithm similar to the HAI algorithm is applied to create rectangular serving areas. Unlike the Sprint analysis, which relied on specific analysis of certain actual clusters in Nevada, the present analysis is based on a Monte Carlo simulation of a large number of randomly generated customer locations. This analysis does not attempt to evaluate the actual distribution or feeder algorithms used in the HAI model. Instead, it attempts to test the accuracy of the preprocessing algorithms used in converting geocoded and surrogate geocoded customer locations into rectangular serving areas that are used by the model to construct distribution and feeder plant.

I implemented the test using the MapBasic programming environment (a copy of the test's source code is included as Appendix 1 to this memorandum). Each trial of the test is made up of the following steps:

---

[1] MCI ex parte of April 23, 1998, page 1.

[2] Letter from Richard N. Clarke, AT&T, on behalf of HAI proponents AT&T and MCI, to Magalie Roman Salas, FCC, dated February 3, 1998 (CC Docket No. 96-45).

(1) I randomly generate a series of N points. The points all lie within a 18 kft by 18 kft box. The set of points represents a cluster of customer locations that could be served by a single SAI.

(2) I calculate two different measures of dispersion for the randomly generated points. The first measure is the total length of a Star Network. A Star Network is constructed by connecting each point in the cluster directly to the centroid of the cluster. The centroid of the cluster is represented by the average X and Y coordinates of the points. The second measure is the length of the Minimum Spanning Tree (MST).[3] The MST is the shortest possible connect-the-dots type graph that includes every point in the cluster. I do not want to imply that either measure is a good measure of the length of a cluster's distribution plant. Both measures, however, are measures of the dispersion of points.

(3) I apply the HAI algorithm's implications, as I then understood them.[4] I first calculate the area and aspect ratio of the convex hull of the points. I then find the height and width of a rectangular distribution area that has the same area and aspect ratio as the convex hull. Next, I find the dimensions of the individual customer's lots. Like HAI's algorithm, I constrain the dimensions of the customer's lots so that: (1) the area of N lots is equal to the area of the rectangular distribution area, and (2) the height of a lot is twice its width.

Once I have calculated these various dimensions, I "lay out" the lots in rows and columns. I make the first column by stacking lots end-to-end and North-to-South in the rectangle (assuming North-to-South is the long side). I stack lots on the first column until there is no more room in the rectangle; I allow the last lot to overflow the rectangle. I then add as few additional columns as possible. I only add columns until the total number of lots is just greater than or equal to the number of customers, N.

Once I have laid out the lots, I allocate customers as if they were in density zone 1 or

---

[3] For information about Minimum Spanning Trees and an algorithm used to find them, see Prim, R.C., November 1957, "Shortest Connection Matrix Network and Some Generalizations," *Bell System Technical Journal*: 36, 1389-1401.

[4] After further conversations with various HAI proponents, I am convinced that my test actually overstates the amount of dispersion implied by the HAI algorithm. The HAI algorithm does not actually determine the location of all individual lots, like I have done. The algorithm only determines the location of four lots, one in each corner of the rectangular distribution area. The HAI model then builds plant as if the customers were all located within the boundaries implied by these four lots. There is an internal difficulty here, because N lots of the prescribed dimensions often don't fit within those boundaries. I've ignored this problem and allowed the lots to overflow the boundaries. Because of this, my estimates of the dispersion implied by the HAI algorithm may be too high.

2.[5] I locate the first customer in the most South-Western lot, half-way along the lot's frontage (or width) and 150 feet South of it's Northern border. I add additional customers, each one lot width to the East, until there are no more lots in the row. I start the second row 300 feet North of the first row. I start the third row two lot heights North of the first row. I continue this pattern until I have located N customers, then I stop. If there are any unpopulated lots, and there typically are, they are in the Eastern part of the most Northern row.

If there are an odd number of North/South lots, I move the customers in the most Northern row from the top to the bottom of their lots.

(4) I calculate the length of the Star Network and of the MST for these new HAI points.

Please note that what I have described above is only a test of the accuracy of the customer locations used in the preprocessing stages of the HAI algorithm. It is not a test of the adequacy of the distribution plant subsequently built by the HAI model given those locations.

At this time, I have performed 2440 trials of this experiment. I have 15 trials for each N in [5,100] and 10 trials for each N in [101,200]. I have computed the error rate for each trial, a negative error meaning that HAI underestimated the dispersion. For example, in trials with N equal to 25, the HAI algorithm underestimates the length of the Star Network by an average of 15.4 percent. It underestimates the length of the MST by an average of 41.5 percent. These underestimates correspond to error rates of -0.154 and -0.415, respectively.

I have also made kernel smooths of the data. Kernel smoothing is a non-parametric regression technique that allows you to plot an underlying trend line given "noisy" data.[6] For simplicity, I used a uniform kernel to estimate the HAI algorithm's error rate as a function of N. The estimation process is equivalent to using an unweighted five period moving average. Figure 1 is a plot of the kernel smooths. The solid line represents the expected error rate, or bias, in the length of the Star Network, the broken line represents the expected bias in the length of the MST.

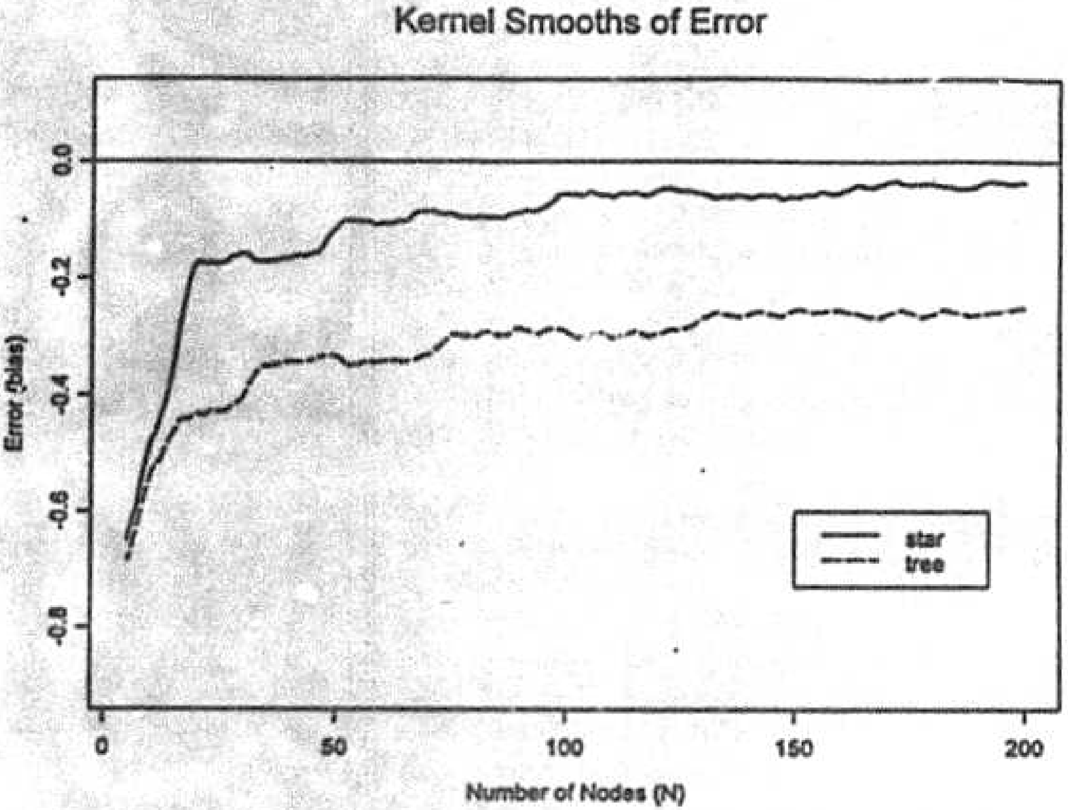
As the graph illustrates, the expected bias is negative over the whole range of N, for both measures. The bias seems to be a function of N, with the size of the bias increasing as N gets smaller. The shape of the curves suggests the following simple parameterization of the relationship between N and the error rate, E:

---

[5] The HAI model defines density zone 1 as an area with 0-5 lines per square mile, and density zone 2 as an area with 6-100 lines per square mile.

[6] For more information about non-parametric regression and smoothing see: Manski, C.F., March 1991, "Regression," *Journal of Economic Literature* XXIX: 34-50, and Härdle, W., 1989, Applied Nonparametric Regression, Cambridge: Cambridge University Press.

Figure 1:



$$E=b_1+b_2\ln(N)+\varepsilon,$$

where b_1 is an intercept parameter, b_2 is a slope parameter, and ε is a random disturbance. Using this parameterization, I've run regressions of the error rates under both measures. In both cases, the relationship between E and N was statistically significant, with p-values essentially equal to zero. The results of the regressions are in the Table 1. Figure 2 shows the curves implied by the estimation.

Figure 2:

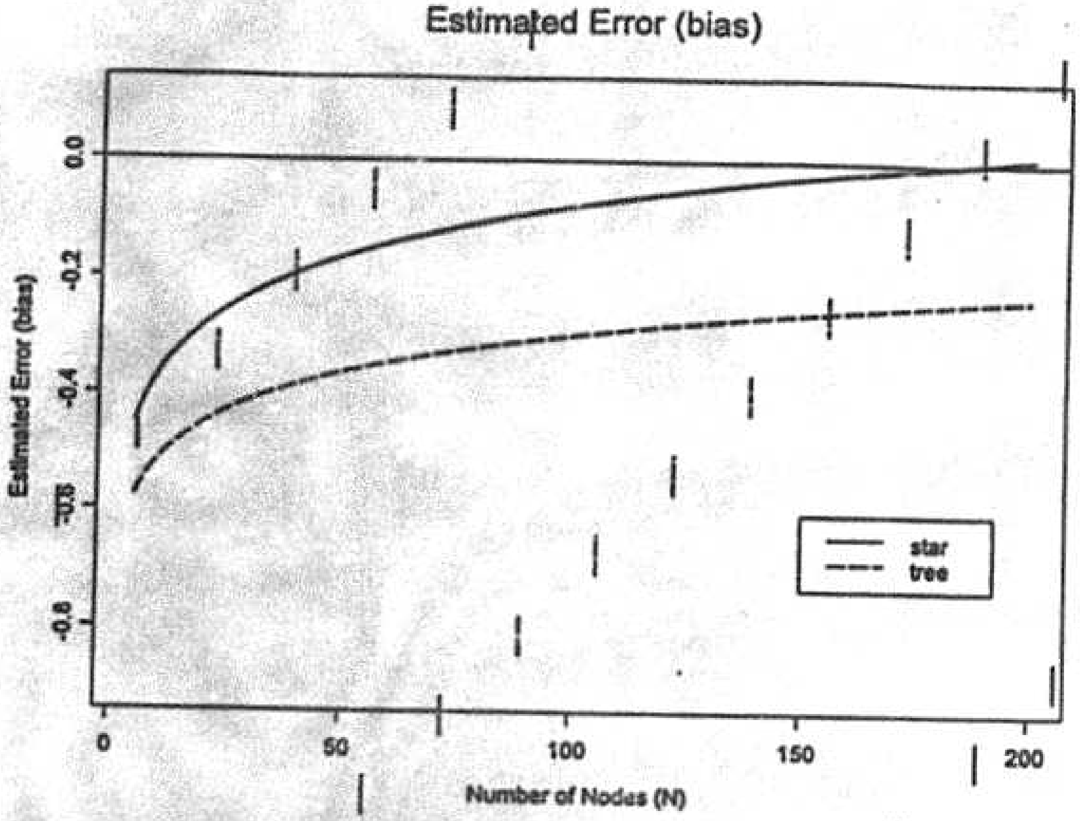

Estimated Error (bias)
Estimated Error (bias)
0.0
-0.2
-0.4
-0.6
-0.8
0
50
100
150
200
Number of Nodes (N)
star
tree

Table 1:

Star	Value	Standard Error	t-stat	Prob (>|t|)
b1	-0.6471	0.0069	-94.0937	0.00
b2	0.1238	0.0016	78.2301	0.00
R-Squared	0.7151			
Residual Standard Error	0.06531			
df	2438			

MST	Value	Standard Error	t-stat	Prob (>|t|)
b1	-0.7268	0.0038	-193.5693	0.00
b2	0.0937	0.0009	108.4149	0.00
R-Squared	0.8282			
Residual Standard Error	0.03566			
df	2438			

Appendix 1: The MapBasic source code for running the test.

I performed my test in two series. The first series consisted of 5 trials for each N in [5,100]. You should be able to replicate this series by seeding the random number generator with -1. The second series consisted of 10 trials for each N in [5,200]. You should be able to replicate this series by seeding the random number generator with -2. Note: you can toggle interactive mode on and off before you compile.

```
'-- May 1, 1998.
'-- This program conducts a Monte-Carlo type test of HAI V5.0's Customer
'-- Location Algorithm, as I understand it.
'--
'-- Jeff Prisbrey
'-- Federal Communications Commission

Include "Mapbasic.Def"

Declare Sub Main
Declare Function Hull(X(),Y() as Float, HullInd(), NoPts, HullCnt as Integer) as Logical
Declare Function Tree(X(),Y() as float, ToNode(), N as integer) as Logical
Declare Function Rnd3(ByVal idum As Integer) As Float

'-- I need these global variables for the random number generator.
Global iff  inext, inextp, ma(55) As Integer


Sub Main
Close All
Dim X(300), Y(300), nX(300), nY(300), rX, rY as Float
Dim HullInd(300). HullCnt, ToNode(300), N as Integer
Dim NSlots, EWlots, Tlots as Integer
Dim i, j, k, t1, t2 as Integer
Dim starD1, starD2, treeD1, treeD2 as float
Dim cenX, cenY, Area1, AspectR, H, W, a, b as Float
Dim Result as Logical
Dim RegionObj as Object
Dim RectObj as Object

'-- Initialize Knuth's random number generator.
H = Rnd3(-2)
H = 0.0

'-- Create a place to store the results.
```

```
Create Table Results
  (Trial integer,
    NumNodes integer,
    ETime integer,
    StarDist1 float,
    StarDist2 float,
    TreeDist1 float,
    TreeDist2 float
    )
  File ApplicationDirectory$() + "ResultsB"
  Type DBF
Set Table Results
  FastEdit On
  Undo Off

Set Window Message Position (7,.75) Height 3.5
Set CoordSys Nonearth Units "ft" Bounds (-180000,-180000) (180000,180000)

N = 4
For j = 1 to 1960
 t1 = timer()
 starD1 = 0.0
 starD2 = 0.0
 treeD1 = 0.0
 treeD2 = 0.0

 '-- Create a place for the objects, and a coordinate system for plotting.
 If NumTables() > 1 Then
   If TableInfo(Cluster, Tab_Info_NRows) > 0 Then
     Drop Table Cluster
   End If
 End if
 Create Table Cluster
   (Id Char(10))
   File ApplicationDirectory$() + "Cluster"
 Set Table Cluster
   FastEdit On
   Undo Off
 Create Map For Cluster
   CoordSys Nonearth Units "ft" Bounds (-180000,-180000) (180000,180000)
 Map from Cluster
 Set Map Center (9000,9000) Layer 1 Editable On Selectable On

 '-- Increment N if needed, pick random points.
```

```
'-- Note that the points are within an 18 kft box and could therefore
'-- represent customer locations in a single cluster.
If ((j - 1) MOD 10) = 0 Then
  N = N + 1
End If
For i = 1 to N
  X(i) = Rnd3(1) * 18001
  Y(i) = Rnd3(1) * 18001
  Create Point(X(i), Y(i))
Next
Set Map Zoom Entire Layer 1

'-- Find the points which define the convex hull of the cluster.
'-- Create the convex hull and insert it into the table
Result = Hull(X, Y, HullInd, N, HullCnt)
Create Region Into Variable RegionObj 0 Brush (1,0,0)
For i = 1 to HullCnt
 Alter Object RegionObj Node Add (X(HullInd(i)),Y(HullInd(i)))
Next
Insert Into Cluster (obj) Values (RegionObj)

'-- Find the centroid of the cluster.
cenX = 0.0
cenY = 0.0
For i = 1 to N
  cenX = cenX + X(i)
  cenY = cenY + Y(i)
Next
cenX = cenX / N
cenY = cenY / N

'-- Find the distance in the initial "Star" network.
'-- Find the distance in the inital minimum spanning tree.
Result = Tree(X, Y, ToNode, N)
starD1 = Distance(cenX, cenY, X(1), Y(1), "ft")
treeD1 = 0.0
For i = 2 to N
  treeD1 = treeD1 + Distance(X(i),Y(i),X(ToNode(i)),Y(ToNode(i)),"ft")
  Create Line (X(i) ,Y(i)) (X(ToNode(i)) , Y(ToNode(i)))
    Pen Makepen(1,2,RGB(255,0,0))
  starD1 = starD1 + Distance(cenX, cenY, X(i), Y(i), "ft")
Next

'-- HAI treats areas less than 0.03 square miles as high-rise buildings.
```

```
'-- I'll exclude consideration of these for now.
If Area(RegionObj, "sq mi") > 0.03 then

  '-- Compute the area and aspect ratio for the representative rectangle.
  Area1 = Area(RegionObj, "sq ft")
  AspectR = Distance(ObjectGeography(RegionObj,OBJ_GEO_MINX),
                ObjectGeography(RegionObj,OBJ_GEO_MINY),
                ObjectGeography(RegionObj,OBJ_GEO_MINX),
                ObjectGeography(RegionObj,OBJ_GEO_MAXY),
                "ft") /
          Distance(ObjectGeography(RegionObj,OBJ_GEO_MINX),
                ObjectGeography(RegionObj,OBJ_GEO_MINY),
                ObjectGeography(RegionObj,OBJ_GEO_MAXX),
                ObjectGeography(RegionObj,OBJ_GEO_MINY),
                "ft")

  '-- For simplicity, always make the rectangle taller than it is wide.
  If AspectR < 1 then
    AspectR = 1 / AspectR
  End If

  '-- The dimensions of the rectangular region.
  W = Sqr(Area1 / AspectR)
  H = W * AspectR

  '-- For plotting purposes, this is the lower corner of the rectangular region.
  rX =  ObjectGeography(RegionObj,OBJ_GEO_MINX) - W * 3 / 2
  rY =  ObjectGeography(RegionObj,OBJ_GEO_MINY)

  '-- Create the rectangular region and insert it into the table.
  Create Region Into Variable RectObj 0 Brush (1,0,0)
    Pen Makepen(1,2,RGB(0,0,255))
  Alter Object RectObj Node Add (rX,rY)
  Alter Object RectObj Node Add (rX,rY+H)
  Alter Object RectObj Node Add (rX+W,rY+H)
  Alter Object RectObj Node Add (rX+W,rY)
  Insert Into Cluster (obj) Values (RectObj)

  '-- a is the width of a lot, b is the depth.
  a = Sqr(Area1/(2*N))
  b = 2*a

  '-- Determine the number of NS and EW lots
  NSlots = (H \ b) + 1
```

```
 EWlots = ((N - 1) \ NSlots) + 1
 Tlots = NSlots * EWlots

 '-- Plot the lot regions.
 For i = 0 to EWlots
   Create Line (rX + i*a, rY) (rX + i*a, rY + NSlots*b)
     Pen Makepen(1,2,RGB(0,255,0))
 Next
 For i = 0 to NSlots
   Create Line (rX , rY+ i*b) (rX + EWlots*a, rY + i*b)
     Pen Makepen(1,2,RGB(0,255,0))
 Next


'-- Populate the lots, calculate customer locations.
'-- Start at location 1 and work your way east.  When you
'-- get to the end of the lots, move up a row.
'-- The variable k follows the rows.
nX(1) = rX + a / 2
nY(1) = rY + (b - 150)
For i = 2 to N
  nX(i) = nX(1) + a * ((i-1) MOD EWlots)
  k = (i-1) \ EWlots
  nY(i) = nY(1) + (k MOD 2) * 300 + ((k \ 2) * 2) * b
Next '-- i

'-- If the number of rows is odd, bring the locations in the top row
'-- down to the bottom part of their lots.
If (NSlots MOD 2) = 1 then
  For i = Tlots to (Tlots - EWlots + 1) Step -1
    nY(i) = nY(i) - b + 300
  Next
End If

'-- Find the centroid of the new points.
cenX = 0.0
cenY = 0.0
For i = 1 to N
  Create Point (nX(i), nY(i))
  cenX = cenX + nX(i)
  cenY = cenY + nY(i)
Next
cenX = cenX / N
cenY = cenY / N
```

```
    Set Map Zoom Entire Layer 1

    '-- Find the distance in the new "Star" network.
    '-- Find the distance in the new minimum spanning tree.
    Result = Tree(nX, nY, ToNode, N)
    starD2 = Distance(cenX, cenY, nX(1), nY(1), "ft")
    treeD2 = 0.0
    For i = 2 to N
      treeD2 = treeD2 + Distance(nX(i),nY(i),nX(ToNode(i)),nY(ToNode(i)),"ft")
      Create Line (nX(i) , nY(i)) (nX(ToNode(i)) , nY(ToNode(i)))
        Pen Makepen(1,2,RGB(255,0,0))
     starD2 = starD2 + Distance(cenX, cenY, nX(i), nY(i), "ft")
    Next

    '-- Write the results to the screen.
    Print Chr$(12)
    Print "Trial No. " + Str$(j)
    Print "Time for Trial: " + Str$(Timer() - t1)
    print "N: " + Str$(N)
    print " "
    print "starD1: " + Str$(starD1)
    print "starD2: " + Str$(starD2)
    print "%error: " + Str$(100*(starD2 - starD1)/starD1)
    print " "
    print "treeD1: " + Str$(treeD1)
    print "treeD2: " + Str$(treeD2)
    print "%error: " + Str$(100*(treeD2 - treeD1)/treeD1)

End If '-- Not a high-rise.

'-- Write the results to the Results table.
t2 = Timer() - t1
Insert Into Results (Trial,NumNodes,ETime,StarDist1,StarDist2,TreeDist1,TreeDist2)
  Values (j,N,t2,starD1,starD2,treeD1,treeD2)

'-- Switch the > or < sign to toggle interactive mode.
If N > 0 then
  Dialog Title " Do It Again? " Width 100 Height 50 Position 50,540
    Control OKButton Title "Again"
    Control CancelButton Title "Exit"
  If not CommandInfo(CMD_INFO_DLG_OK) Then
    Exit For
  End If
End If
```

```
Next '-- j
Close All
End Sub '-- Main


Function Hull(X(),Y() as Float, HullInd(), NoPts, HullCnt as Integer) as Logical
  '-- This function is taken from code published on the internet by The MapTools Company.
  '-- It's an implementation of Graham's Algorithm for finding the convex hull of a
  '-- set of points.

  ' The Hull function returns "False" if NoPts < 3 or for duplicate points
  '    or all the points on a line

 Dim Dist, T, Angle, D2MN, D2IM as Float
 Dim I, J, M, N, AngleInd as Integer
 Hull = False
 If NoPts < 3 Then Exit Function End If

  ' Find the pair M,N with the greatest separation
 Dist = 0
 For I = 1 to NoPts
   For J = I+1 to NoPts
    T = (X(J)-X(I))^2 + (Y(J)-Y(I))^2
    If T > Dist Then
      M = I   N = J   Dist = T
    End If
   Next
 Next
 If Dist = 0 Then Exit Function End If

  ' Find the rightmost point from the line M to N
 Angle = -1
 AngleInd = M
 For J = 1 to 2
   D2MN = ( (X(N)-X(M))^2+(Y(N)-Y(M))^2 )
   If D2MN = 0 Then Exit Function End If
   For I = 1 to NoPts
     If I <> M and I <> N Then
       D2IM = (X(I)-X(M))^2+(Y(I)-Y(M))^2
       If D2IM = 0 Then  Exit Function  End If
       T = ((X(I)-X(M))*(Y(N)-Y(M)) - (X(N)-X(M))*(Y(I)-Y(M)))/
         Sqr(D2IM * D2MN)
       If (T > Angle) Then Angle = T AngleInd = I End If
       If (T = Angle and (X(I)-X(M))^2 + (Y(I)-Y(M))^2 <
```

```
       (X(AngleInd)-X(M))^2 + (Y(AngleInd)-Y(M))^2) Then
         AngleInd = I End If
     End If
    Next
    ' Test to make sure some point is to the right,
    '    otherwise exchange M and N and do again
    If Angle > 0 Then Exit For End If
    If J = 2 then Exit Function End If
    AngleInd = N    N = M    M = AngleInd
   Next
   HullInd(1) = M    HullInd(2) = AngleInd   HullCnt = 2
   N = M   M = AngleInd

   ' Find the point at the greatest clockwise angle from current Hull edge
   Do
    Angle = 1   AngleInd = M
    D2MN = ( (X(N)-X(M))^2+(Y(N)-Y(M))^2 )
    For I = 1 to NoPts
     If I <> M and I <> N Then
      D2IM = (X(I)-X(M))^2+(Y(I)-Y(M))^2
      If D2IM = 0 Then Exit Function End If
      T = ((X(I)-X(M))*(X(N)-X(M)) + (Y(I)-Y(M))*(Y(N)-Y(M)))/
       Sqr(D2IM * D2MN)
      If (T < Angle) Then
       Angle = T  AngleInd = I  End If
      If (T = Angle and (X(I)-X(M))^2 + (Y(I)-Y(M))^2 <
       (X(AngleInd)-X(M))^2 + (Y(AngleInd)-Y(M))^2) Then
         AngleInd = I End If
     End If
    Next
    HullCnt = HullCnt + 1   HullInd(HullCnt) = AngleInd
    N = M   M = AngleInd
   Loop Until HullInd(HullCnt) = HullInd(1)
   Hull = True
End Function


Function Tree(X(), Y() as float, ToNode(), N as integer) as Logical
  '-- This is an implementation of Prim's minimum spanning tree algorithm.

  Dim inTree(300) as logical
  Dim minDist(300) as float
  Dim d, minD as float
  Dim i, j, k, m as integer
```

```
    For i = 1 to N
      inTree(i) = False
      ToNode(i) = -1
      minDist(i) = 1.79769313486232E+307
    Next
    inTree(1) = True

    j = 1
    For i = 2 to N
      minD = 1.79769313486232E+307
      For k = 2 to N
        If (inTree(k) = False) then
          d = Distance(X(k),Y(k),X(j),Y(j),"ft")
          If d < minDist(k) then
            minDist(k) = d
            ToNode(k) = j
          End If
          If minDist(k) < minD then
            minD = minDist(k)
            m = k
          End If
        End If '-- inTree
      Next '-- k
      inTree(m) = True
      j = m
  Next '-- i

  Tree = True
End Function


Function Rnd3(ByVal idum As Integer) As Float
'-- Subtractive Random Number Generator due to Knuth.  From Numerical Recipes.
'-- This is better than an LCR generator for doing Monte-Carlo type stuff.

Dim i, ii, k As Integer
Dim mj, mk As Integer

If (idum < 0) Or (iff = 0) Then
  iff = 1
  mj = 161803398 - Abs(idum)
  mj = mj Mod 1000000000
  ma(55) = mj
```

```
    mk = 1
    For i = 1 To 54
      ii = (21 * i) Mod 55
      ma(ii) = mk
      mk = mj - mk
      If (mk < 0) Then mk = mk + 1000000000 End If
      mj = ma(ii)
    Next
    For k = 1 To 4
      For i = 1 To 55
        ma(i) = ma(i) - ma(1 + (i + 30) Mod 55)
        If (ma(i) < 0) Then ma(i) = ma(i) + 1000000000 End If
      Next
    Next
    inext = 0
    inextp = 31
    idum = 1
  End If

  inext = inext + 1
  If (inext = 56) Then inext = 1 End If
  inextp = inextp + 1
  If (inextp = 56) Then inextp = 1  End If
  mj = ma(inext) - ma(inextp)
  If (mj < 0) Then mj = mj + 1000000000 End If
  ma(inext) = mj
  Rnd3 = mj * 1.0 / 1000000000.0

  End Function
```

Sprint

Jay C. Keithley
Vice President

Law & External Affairs
1850 M Street, NW, Suite 1100
Washington, DC 20036
Voice 202 828 7453
Fax 202 296 3469
jay.c.keithley@mail.sprint.com

EX PARTE

May 19, 1998

Ms. Magalie Roman Salas
Secretary
Federal Communications Commission
1919 M Street, N.W. Room 222
Washington, D.C. 20554

RE: In the Matter of Federal-State Joint Board on Universal Service, CC Docket No. 96-45 and Forward-Looking Mechanism for Non-Rural LECs, CC Docket No. 97-160

Dear Ms. Salas:

The attached was today provided to the Common Carrier Bureau staff listed below in respopnse to the HAI Model analysis of Jeff Prisby dated May 13, 1998.

Sprint requests that this information be made a part of the record in this matter. Four copies of this letter, in accordance with Section 1.1206(a)(1), are provided for this purpose. If you have any questions, please feel free to call.

Sincerely,

Jay C Keithley

Jay C. Keithley

Attachment

cc:

Brad Wimmer	Brian Clopton
Chuck Keller	Natalie Wales
Don Stockdale	Jim Schlichting
Craig Brown	Bob Loube
Richard Smith	Lisa Gelb

Sprint Response to Analysis of Jeffrey Prisbrey

In his document dated May 13, 1998, Jeffrey Prisbrey presents evidence regarding measures of customer dispersion similar to those modeled in the Hatfield Model 5.0a (a.k.a. HAI Model). Specifically, as stated in his document, Mr. Prisbrey provides a "measure of dispersion of customer locations both before and after an algorithm similar to the HAI algorithm is applied to create rectangular serving areas."

The spreadsheet that accompanies Mr. Prisbrey's document presents the following information and conclusion: **In general, customer dispersion is altered, often dramatically, when "an algorithm similar to the HAI algorithm" is applied to customer locations.** The bias of this impact is uniformly negative.

The result of this bias produced by the HAI Model (and its preprocessing) is to understate customer dispersion and, all else held constant, understate the cost of providing service to customers by underbuilding the network, specifically with regard to distribution plant. This finding is consistent with previous findings that Sprint has presented to the FCC.

Specifically, Sprint has shown numerous cases where the amount of distribution cable built by the HAI Model for individual clusters falls far short of the amount of cable actually required to connect customer locations in those clusters. One driving factor behind this underbuilding is the distortion of customer location that occurs in the model's preprocessing (the exact occurrence described by Mr. Prisbrey). Further distortion takes place in the HAI Distribution Module itself. The result of this underbuilding is that the Hatfield Model 5.0a does not build a functioning telephone network in many of the areas that are of greatest concern for universal service.

Recently, Sprint was provided an opportunity to examine additional data for the state of Nevada to determine the extent of this underbuilding problem. As stated in the Prisbrey document, the HAI Sponsors have claimed that this situation is rare. Mr. Prisbrey's analysis strongly suggests otherwise, and provides evidence that the problem is systematic rather than random.

On the following pages, Sprint presents summary evidence that provides further support for Mr. Prisbrey's analysis. The following pages contain no protected or proprietary material.

Mr. Prisbrey found, on average, that the *number of customer locations per cluster* was related to the amount of distortion the HAI produced. For example, distortion was larger (in percentage terms) for smaller clusters (N=5) than for larger clusters (N=40). This is shown below.

Number of Customer Locations per Cluster	Average Distortion from Star Network	Average Distortion from Minimum Spanning Tree
N=5	71.2%	73.2%
N=10	47.6%	55.5%
N=15	36.0%	44.0%
N=25	15.4%	41.5%
N=40	14.8%	32.8%

These are consistent with the results Sprint found in its own analysis.

The table below lists the amount of underbuilding (as a result of this customer location distortion) that the Hatfield Model exhibited for Nevada with regard to 2 measures: the minimum spanning tree for the cluster, and the diagonal distance of the cluster's minimum bounding rectangle.

(For example, assume the minimum spanning tree of a cluster is 5000 feet, and the diagonal of that cluster's minimum bounding rectangle is 3000 feet. If the HAI Model builds a total of 1000 feet of distribution for that cluster, the first column below would show an underbuild of 300% (3000/1000) and the second column would show an underbuild of 500% (5000/1000).

Number of Customer Locations per Cluster (Main Clusters Only)	Average Underbuild from Diagonal of Minimum Bounding Rectangle	Average Underbuild from Minimum Spanning Tree for Cluster
N=5	1,074%	1,911%
N=10	406%	511%
N=15	255%	358%
N=20	121%	225%
N=25-30	61%	141%

[It should be noted that the vast majority of the smallest clusters (N=5 or N=10) all fall within the two lowest density zones for the state of Nevada and these two density zones account for over 90% of the universal service support for the state. Therefore, this distortion and underbuilding is greatest in those specific areas that are of greatest concern for USF purposes.]

As the table shows, the relationship between size of the cluster (N) and location distortion is exhibited in the relationship between size of the cluster and amount of underbuilding. It should also be noted that the percentages are dramatically higher in the table based on actual Nevada data. The reason for this is straightforward, and is outlined below:

As stated in Sprint's ex parte filing of April 17, 1998, the HAI model underbuilds as a result of three separate effects

1. The conversion of the original polygon to a reduced rectangle (this is the only portion captured by Prisbrey's analysis), which reduces customer dispersion
2. The practice of not building to the outside of the perimeter lots, which reduces customer dispersion even further, and
3. The assumption that lots are shaped and situated a specific way, and that all customers live within 150 feet of the front of the lot.

The combined impact of these three effects results in the dramatic underbuilding that Sprint has documented since April. Furthermore, evidence of this underbuilding has appeared in the rural areas of every state we have investigated. This lends further support for Prisbrey's conclusion that this is not a random or rare occurrence, but a systematic effect.

The claim regarding the frequency of these occurrences is well supported in Prisbrey's document. The figures shown in the table below were obtained using the spreadsheet provided with that document.

Number of Clusters in Prisbrey analysis	Number of Clusters where Hatfield Dispersion was LESS than Original Disperion as measured by Star Network	Number of Clusters where Hatfield Dispersion was LESS than Original Dispersion as measured by Minimum Spanning Tree.
2440	2399 (98%)	2400 (100%)

This fact is also supported by Sprint's evaluation of the actual data from Nevada.

Number of Main Clusters having 200 Lines or Less (maximum N in Prisbrey analysis)	Number of these Clusters where Hatfield Distribution was LESS than the distance of the diagonal of the cluster's minimum bounding rectangle.	Number of these Clusters where Hatfield Distribution was LESS than the distance of the cluster's minimum spanning tree.
586	449 (76%)	566 (96%)

The distortion cited by Mr. Prisbrey and its impact on the network and costs that the HAI produces is not rare, it is not random, and it is not *de minimis*.

Based on this evidence and other evidence presented to date, the following is clear: **The Hatfield (or HAI) Model 5.0a in its current form cannot be used as a costing methodology for calculating explicit USF.**

Sprint respectfully submits the following Model Recommendation:

Sprint believes that the BCPM produces a substantially more accurate estimate of distribution plant distances and associated costs that the HAI Model. However, Sprint also recognizes that an even more accurate estimate of distribution plant and costs, especially in less populated rural areas, could be produced using actual (i.e. geocoded) customer locations. Sprint also agrees with the Commission staff's determination to obtain that customer location data. However, *the key to this additional accuracy lies in the model (any model) actually using the customer location: building plant to actual locations, and maintaining relative distances between locations.*

In order to finalize a forward looking cost model by August 8th, Sprint urges the Commission to take the following steps:

1. The Commission needs to resolve the outstanding network design and technical parameter issue. These include, for example, the maximum copper loop length (the BCPM Sponsors recommend use of a 12,000-foot limit, while the HAI Sponsors recommend a limit of 18,000 feet), and the method for serving very sparsely situated customers (the BCPM serves these customers through DLCs, while the HAI uses T1 repeaters and remote terminals.) These issues are well articulated in the various comments and ex parte submissions of interested parties, and are ripe for Commission decision. The Commission should note the possibility that this resolution might include incorporating portions of each model into the Commission's final model (see point 2 immediately below).
2. Most importantly, the Commission needs to take "ownership" of the modeling effort. Based on the models submitted for its consideration, and the resolution of the network design and technical parameter issue noted above, the Commission should take responsibility for finalizing its cost model. Sprint believes it is no longer productive to continue the development of competing, privately funded cost models. Only by taking ownership of the model development process, and using the work that has already been done by the model sponsors, can the Commission hope to meet its August 8th deadline.
3. The Commission should develop a plan and timeline for obtaining geocoded data from all LECs. Sprint believes it would be reasonable to require that such data be produced by mid-2000, and incorporated into the model for cost estimation purposes by January, 2001.

Exhibit BKS-4

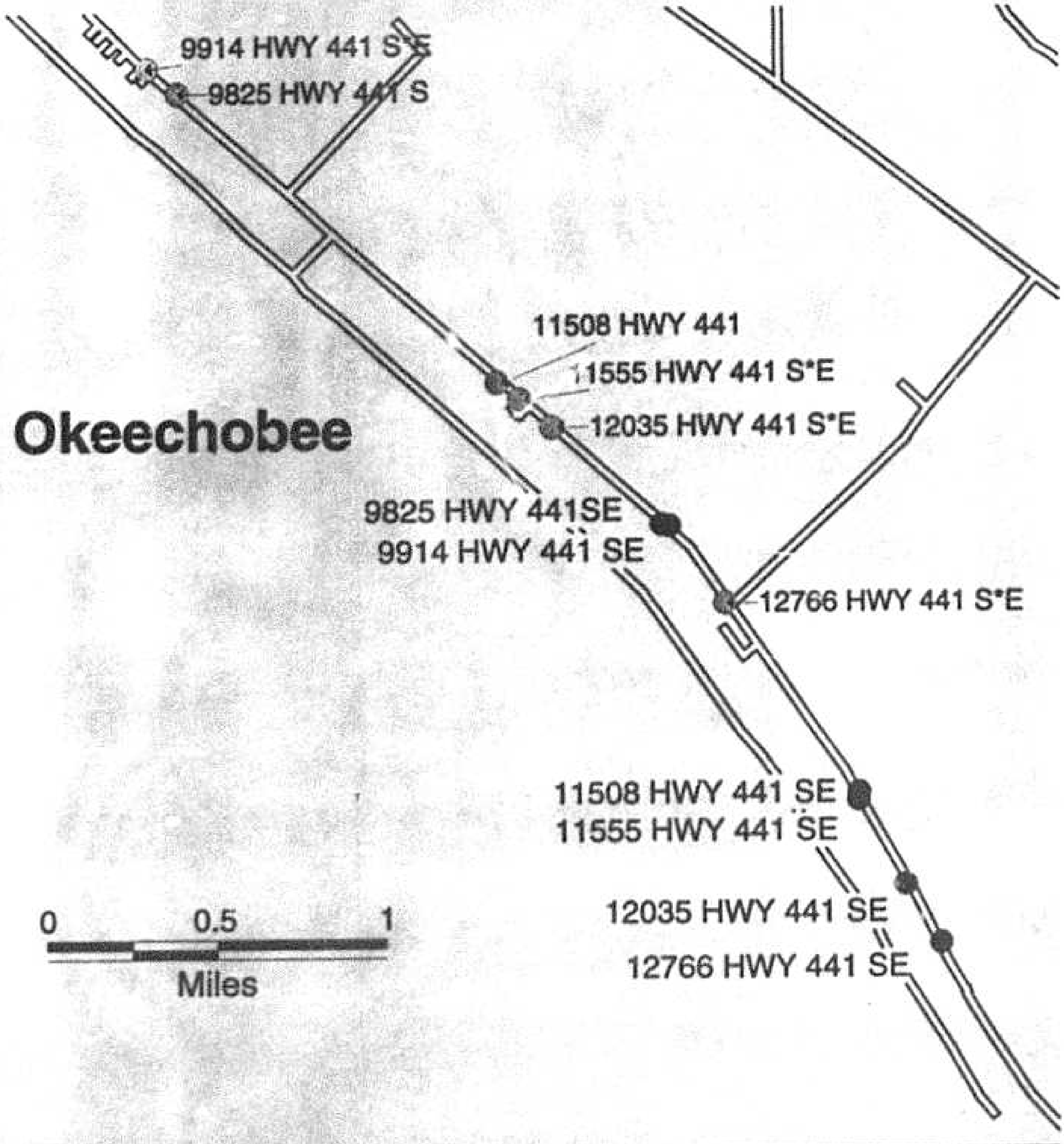

9914 HWY 441 S*E
9825 HWY 441 S
11508 HWY 441
11555 HWY 441 S*E
Okeechobee
12035 HWY 441 S*E
9825 HWY 441SE
9914 HWY 441 SE
12766 HWY 441 S*E
11508 HWY 441 SE
11555 HWY 441 SE
0
0.5
1
Miles
12035 HWY 441 SE
12766 HWY 441 SE

Exhibit BKS-5

RECEIVED
AUG 28 1998
FEDERAL COMMUNICATIONS COMMISSION
OFFICE OF THE SECRETARY

Before the
FEDERAL COMMUNICATIONS COMMISSION
Washington, DC 20554

In the Matter of	)	
	)	
Federal-State Joint Board on Universal Service	)	CC Docket No. 96-45
	)	
	)	
Forward-Looking Mechanism For High Cost Support For Non-Rural LECs	)	CC Docket No. 97-160 (DA-98-1587)

JOINT COMMENTS OF BELLSOUTH TELECOMMUNICATIONS, INC., U S WEST, INC., AND SPRINT CORPORATION TO COMMON CARRIER BUREAU REQUEST FOR COMMENT ON MODEL PLATFORM DEVELOPMENT

BELLSOUTH TELECOMMUNICATIONS, INC.

M. Robert Sutherland
Richard M. Sbaratta
Suite 1700
1155 Peachtree Street, N. E.
Atlanta, GA 30309-3610
(404) 249-3390

Their Attorneys

August 28, 1998

U S. WEST, INC.

Robert B. McKenna
Suite 700
1020 19th Street, N. W.
Washington, DC 20036
(303) 672-2798

Of Counsel. Dan L. Poole

SPRINT CORPORATION

Jay C. Keithley
Sandra K. Williams
1850 M Street
Suite 1110
Washington, D. C. 20036
(202) 828-7453

Attorneys for Sprint Local Telephone Companies

Before the
Federal Communications Commission
Washington, D.C. 20554

In the Matter of	)	
	)	
Federal-State Joint Board on Universal Service	))	CC Docket No. 96-45
	)	
Forward-Looking Mechanism For High Cost Support For Non-Rural LECs	))	CC Docket No. 97-160 (DA-98-1587)

JOINT COMMENTS OF BELLSOUTH TELECOMMUNICATIONS, INC., U S WEST, INC., AND SPRINT CORPORATION TO COMMON CARRIER BUREAU REQUEST FOR COMMENT ON MODEL PLATFORM DEVELOPMENT

BellSouth Telecommunications, Inc., Sprint Corporation and US West, Inc. (hereinafter "Joint Sponsors"), joint sponsors of the Benchmark Cost Proxy Model ("BCPM"), hereby submit the following comments in response to the Public Notice released on August 7, 1998.[1]

The Public Notice affords parties an opportunity to update their comments regarding the forward-looking economic cost model platforms that have already been filed with the Commission. The Joint Sponsors welcome the opportunity to provide the Commission with their response to issues relating to the cost model platforms set forth in the Public Notice.

To facilitate the Commission's review, the Joint Sponsors provide Attachment A which provide their comments relating to the issues as outlined in the Public Notice seeking comment about the utilization of customer location data, methods of grouping customers, the design of distribution and feeder plant, and reaction to the current synthesized version of the HCPM.

---

[1] "Common Carrier Bureau Seeks Comment on Model Platform Development," Public Notice, DA 98-1587, released August 7, 1998 (hereinafter "Public Notice").

Respectfully submitted,

BELLSOUTH TELECOMMUNICATIONS, INC.

By:
M. Robert Sutherland
Richard M. Sbaratta

Suite 1700
1155 Peachtree Street, N. E.
Atlanta, GA 30309-3610
(404) 249-3390

Their Attorneys

U S. WEST, INC.

By:
Robert B. McKenna

Suite 700
1020 19th Street, N. W.
Washington, DC 20036
(303) 672-2798

Of Counsel, Dan L. Poole

SPRINT CORPORATION

By: Jay C. Keithley ([illegible])
Jay C. Keithley
Sandra K. Williams

1850 M Street
Suite 1110
Washington, D. C. 20036
(202) 828-7453

Attorneys for Sprint Corporation

August 28, 1998

Attachment A

The Public Notice crafted its questions and search for comment under headings that dealt with utilization of customer location data, methods of grouping customers, the design of distribution and feeder plant, and reaction to the current synthesized version of the HCPM. The comments are organized in the same manner.

Synthesis Concept

The Joint Sponsors are not at all opposed to the concept of synthesizing a cost proxy model platform utilizing different elements from various models. The concept is workable. In fact, to the extent that the better (or best) practice for the different modules is utilized in the synthesized product, that product should be superior to any of its component-producing progenitors. The Joint Sponsors do have a number of concerns, however, about the process. First, it will be difficult to produce a finished product in the desired timeframe. That is not said to dissuade the Commission from attempting the synthesis effort, but said in an effort to manage expectations and more importantly, in an effort to encourage that accuracy and efficacy not be sacrificed on the alter of expediency. It has taken months and months for the BCPM and HAI to come to where they are today. Admittedly, HCPM and its synthesized byproducts can learn from that process, but it will need to undergo a great deal of public scrutiny before it can be proffered as a finished product. The Joint Sponsors are concerned about any further delay in the implementation of federal high cost fund. In this regard, the Joint Sponsors are ready to assist and will continue to assist the FCC staff in meeting the July 1, 1998 implementation date.

One method to increase the ability to test the HCPM would be to create a linkage between modules of the HCPM and BCPM or between HCPM and HAI. It is possible to "weld" modules from one model to another. By "welding" we mean incorporating the HCPM modules for customer location, clustering, and loop cost generation into the HAI or BCPM models so that the user may edit inputs, process the HCPM, and develop an expanded array of reports with which to analyze its process. The use of "welding" techniques can provide valuable information at this time. Such a welding process is currently underway for the existing models with the HCPM.

Utilization of Customer Location Data

Geocode Data

Customer location data is key to the accuracy of any proxy model[1]. As has been stated in the past, the BCPM Joint Sponsors believe good geocoded data would be beneficial to any proxy model. However, the Joint Sponsors have repeatedly questioned the quality of the currently available geocoded data. Our concerns stem from at least three items. First, currently available geocoded points do not account for unpopulated households. To meet the minimum service requirements of most states, unpopulated households must also have public network connections nearby. Otherwise, it would be impossible for carriers of last resort to meet the requirement to provide service within a certain number of days of the request. Second, geocoding success rates are poorest in rural, high cost areas – exactly those areas where universal service subsidy requirements are the greatest. Third, we need to be assured that the source of the geocoding points and the geocoding process are of high quality. That is not always a certainty today.

In addition to a concern about the use of geocoding as it is possible today, the Joint Sponsors are more concerned about accurately locating those customers that are not geocoded. Since we know that the geocoding will not account for all customers, surrogate points will have to be determined for those households that are not geocoded. The proper determination of surrogate points is vital to the accuracy of the models. To compound this concern, not only is the method of creating surrogate points important, so is the method of mixing successfully geocoded points and surrogate points. If that mixing is done improperly it is likely there will be a biased estimate of the actual plant requirements.

Surrogate Methods

As noted in the Public Notice, there have been many options proposed to determine the surrogate location of those households which cannot be geocoded. The BCPM Joint Sponsors have proposed the use of the road network as a leading indicator for locating the households. It is our belief that not only do people live along the roads, but also telephone plant will, to a great extent, follow roads. This is typically where rights-of-way are located. Therefore, in a model trying to replicate the amount of plant, road data is a useful piece of information. We would add that the placement of the surrogate points should be based upon an assignment of the points to the roads. This would produce an unbiased estimate of the household location in the chosen unit of geography.

To address the possible bias of combining geocoded and surrogate customer location points, one possible method is to use a "breakpoint" below which geocoded data points would not be used. That is, if the success rate of geocoded data was not above a certain point, geocoded data should not be used. In speaking with a number of geographical information system (GIS) experts, it is their opinion that this breakpoint needs to be set relatively high (80%-85%) at the census block level. Unfortunately, there is a dearth of empirical evidence to support that level for a breakpoint. Therefore, we will provide a theoretical argument.

There are many possible approaches to the surrogate location process. The Public Notice discusses a uniform distribution on the perimeter (used by HAI), a distribution on the road network (used by BCPM), or a random distribution. Regardless of which approach is taken, when the need for surrogate customer location points arises, the developer must deal with bias. *How can I create surrogate locations that most fairly represent where customers actually exist without diluting the truly accurate customer locations I already have?* We would like to introduce the concept of "calculated placement". Calculated placement takes into account the known points in determining the unknown points. Let's look first at a uniform placement on a perimeter[2], illustrating a pure uniform placement and a calculated uniform placement.

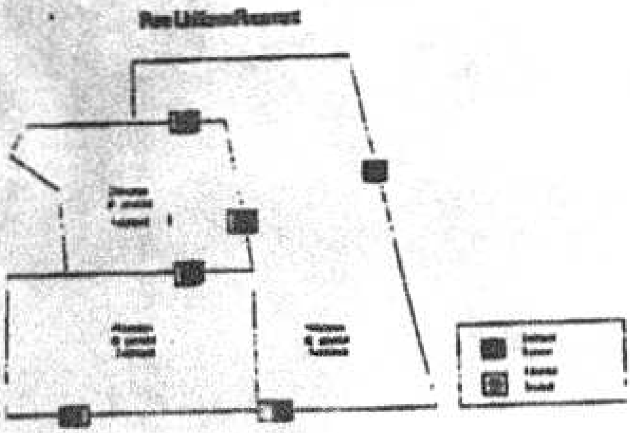
Pure uniform placement can most easily be understood by presenting an example. Assume we are attempting to locate customers in three contiguous census blocks. Using publicly available data, we know 5 customers exist within the first census block, 4 within the second, and

---

[1] Any reference to customers should include both residence and business customers.

[2] Using the HAI model for illustrative purposes here should not be construed as support for that model.

2 within the third. Now assume we are able to successfully geocode 45 % (5) of the customers. creating the need to estimate the locations of the remaining 55% (6) of the customers which we know to exist. Using the *pure uniform placement* process. we would estimate the remaining locations without regard for the successfully located customers - thereby creating the possibility of uniformly placing estimated location points on top of existing points (see diagram below or <u>Attachment C-1</u>).

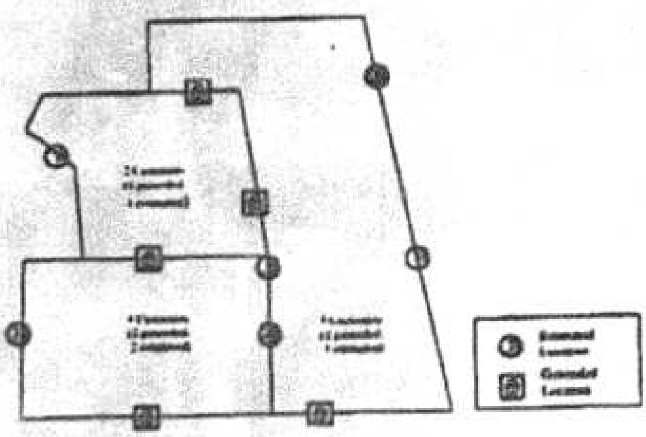


Overlapping the placement of customers may create clusters that do not actually exist. Furthermore, when one considers the vast area that some census blocks cover. the *pure uniform placement* process may greatly understate the cost of supplying universal service.

The logical alternative to the bias created by the *uniform placement* process would be the *calculated uniform placement* process. Using this method, one would supplement the successfully geocoded locations with estimated customer locations that are created <u>acknowledging</u> the existing customer locations.

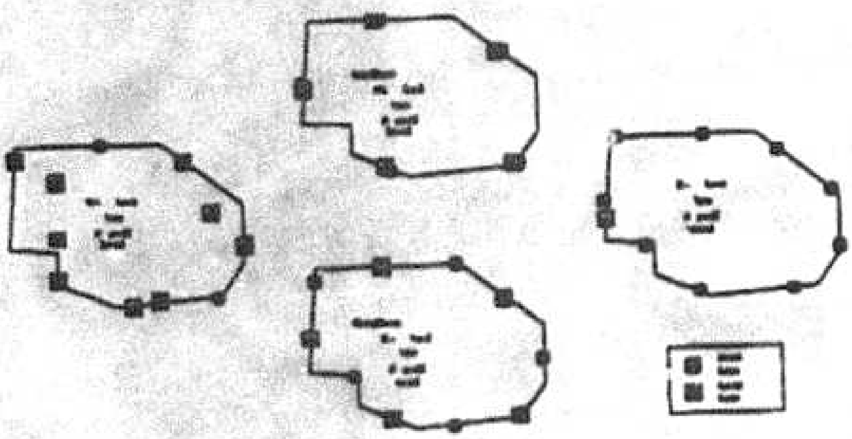
Again assume we are attempting to locate customers in three contiguous census blocks. Using publicly available data, we know 5 customers exist within the first census block, 4 within the second, and 2 within the third. Now assume we are able to successfully geocode 45% (5) of the customers, creating the need to estimate the locations of the remaining 55% (6) of the customers. Using the *calculated uniform placement* process, we identify the successfully geocoded customer location latitudes and longitudes and mark those locations as occupied. The estimated locations are then spaced in an equidistant manner along the remaining perimeter of the census block. Doing so creates unnaturally spaced locations that are not *uniformly* placed at all! This method equally dilutes the impact of successfully geocoded customer locations by incorporating their position into the surrogate location process thereby eliminating any natural clustering or distribution of customers that may truly occur (see diagram below or Attachment C-2).

Calculated Uniform Placement



Thus the combination of geocoded and estimated customer locations creates a bias not present in either process alone. The low success rates of geocoding, particularly in rural areas, allows one to generate data that may appear accurate because it is *based* on successfully geocoded location. However, such locations, when combined with estimated locations, are less representative than estimated location methods alone.

The incorporation of estimated locations in areas already containing extremely high or extremely low geocode success rates, is preferable to doing so in areas containing moderate to low success rates. This is because the chance of creating a skew in customer location by unnaturally spacing or clustering customers is far more likely if one were to combine 50% geocoded locations with 50% estimated locations for example. Consider the diagrams below (see Attachment C-3).



When the successful geocode success rate hits 80% - 85%, the number of estimated locations are so few, that little impact results. However, if a moderate geocode rate is achieved of say, 50%, and the remaining 50% of the customer locations are estimated, the combination of the two placement methods potentially creates one or both of the biases discussed above.

Therefore, we must give special attention to defining the parameters we use to determine when geocoding and estimating customer locations can be used together, and when they cannot. The combination of the two processes can be successfully used when the geocode success rate within a small geographic area (such as a census block, CBG, or wire center) reaches reasonably 85% or higher. This number is obviously not one of precision, but one of reasonable logic. Successfully geocoding the location of 85% of the customers within a wire center generates a set of points that have an impact significant enough to withstand the introduction of the remaining 15% estimated locations without creating clusters or uniform distribution that does not actually exist. In other words, if we know that 85% of the locations are accurate, we can live with a reasonable estimate of the remaining locations. Even if the estimated locations are not truly representative of the actual ones, the bias will not be so great as to significantly change the resulting cost.

Furthermore, by combining customer location methods, an implication of precision is made that may or may not be accurate. If data is presented as being based on geocoded data, the reviewer naturally assumes a high confidence level. That can be quite misleading, however, if the percentage of customers actually geocoded is, in truth, quite low. We should be careful to state the percentage of geocoded locations and the method of surrogate location creation.

Customer Location Data Summary

We understand there may be arguments that any quantity of actual data is better than none. However, we are more concerned about the bias that may be introduced. We would be open to other methods of using geocoded and surrogate points that could address the issue of bias.

Mixing estimated locations with geocoded locations can create clusters with unnatural distributions. Therefore, our recommendation is to use the geocoded locations in the chosen unit of geography (e.g. wire centers, census blocks) where geocode success rates are greater than 85%. In units of geography where geocode success rates are less than 85%, estimated customer locations should be used.

Methods of Grouping Customers

The Commission seeks comment on the relative merits of the HAI Model's clustering algorithm and the Commission staff's clustering algorithm described in the "Test Data" section below.

At this time the BCPM Sponsors are not prepared to offer detailed comments on the specific merits of the HCPM clustering algorithm. Rather, we believe the Commission should focus its attention on the proper use of the results of any clustering algorithm under consideration, and should attempt to avoid the pitfalls of the HAI clustering approach discussed below.

In general, the notion of using standard spatial clustering mechanisms to determine which customer locations should be served as a single carrier serving area (CSA) is perfectly acceptable. The BCPM places customers in microgrids and then *groups* microgrids into CSAs of varying size. As stated in the BCPM Model Methodology (page 28), "Modeling grids that vary in size is tantamount to allowing clusters of customers associated with a particular CSA to vary in density and dispersion." Some parties have raised objections to the BCPM approach, complaining that it arbitrarily separates customers who might otherwise be served together. It is important to note that *this is a characteristic of any clustering methodology*, particularly that used by the HAI Model which can be accurately characterized as an agglomerative technique. The results of HAI's agglomerative approach are highly dependent on the *starting point* from which the rastarized cells are aggregated. If the starting point is changed, the criteria that constrain cluster size and shape will result in a different set of clusters being produced. Customers that would be "served together" under one starting point are "broken apart" under another starting point. Hence, the "cookie cutter" descriptor that the HAI Sponsors have chosen to describe the BCPM, ironically, much more accurately applies to the HAI Model itself!

If this Commission decides the concept of clustering is a reasonable means of determining CSAs, it is important to avoid the subjectivity built into the HAI approach. If an agglomerative approach is used, it is necessary to examine the clusters produced by all starting points and establish reasonable criteria for choosing one final set over another.

Most importantly, it is essential that the locations of the points within clusters be maintained, not discarded, as the clusters are used as serving areas in any proxy model. These parameters include relative *distance* between and *dispersion* among points. The BCPM maintains customer dispersion by dividing CSAs (grids) into quadrants when placing customers. The HAI Model discards existing dispersion information and uniformly distributes within clusters. This is a significant point that cannot be overstated. Any clustering algorithm is designed to reveal groupings based on relative locations. To subsequently ignore the relative locations that create the clusters in the first place is pure folly, and guarantees incorrect results.

Design of Distribution and Feeder Plant

Comparison of Proxy Model Feeder Approaches

The feeder portion of the proxy models have evolved over time. Early on, a simplistic north-south-east-west (NSEW) routing of the main feeder was built with separate subfeeders directed to every census block group (CBG). In subsequent releases, improvements to the feeder algorithms have been made. These improvements ranged from routing the main feeder at an angle to sharing of the sub-feeder. The HCPM implements a further proposed improvement by utilizing minimum spanning trees.

We will try to provide a comparative analysis of the current approaches.

HAI 5.0a Feeder Design

HAI uses two distinct methods to lay out feeder plant. In the first method, feeder emanates from the central office in four cardinal compass point directions. Subfeeder extends from the feeder at right angles to serve the main clusters. This is the default feeder design and is essentially the same as that used in earlier versions of the model. The second option, which the user can select, allows the user to indicate whether the feeder should be pointed ("steered") towards the preponderance of main clusters within the wire center quadrant that the feeder is serving. The model applies a route-to-air multiplier to adjust the feeder distance if this option is chosen. This multiplier is a user-adjustable input.

BCPM 3.1 Feeder Design

BCPM also uses two distinct methods in determining feeder layout. These methods are in the preprocessing module and result in the optimal method being chosen for each central office.

The preprocessing module shows the feeder emanating from each central office in four cardinal compass point directions for 10,000'. After 10,000', feeder is either "pointed" toward the population centroid of the wire center quadrant or is "split" and pointed toward the population centroid of 1/2 of the wire center quadrant. The rationale is that for the first 10,000', feeder routes likely follow roads which are typically oriented north, south, east, and west while in town. Out of town, roads are more likely to diverge from this orientation and feeder routes can be more directly targeted toward population centers.

Subfeeder then extends either vertically or horizontally from the feeder to serve BCPM's serving areas (ultimate grids). Subfeeder is shared, where appropriate, between serving areas.

The BCPM preprocessing logic tests whether a pointed (split) feeder yields a shorter total feeder distance (including subfeeder) than cardinal routing. If not, cardinal routing is used.

HCPM Feeder Design

HCPM has similarities to both the HAI and BCPM. The HCPM has 4 main feeder routes emanating from the central office in the four cardinal compass point directions. From this main feeder, junction points are marked. The HCPM then goes through each FDI point to determine where to route. The determination takes into account the cost of structure and the cost of material for the route under consideration. The routes analyzed are those of the junction point on the main feeder and the previously analyzed and routed FDIs in proximity. Using this minimum spanning tree approach that minimizes total cost, a modified "pine-tree" approach is constructed to connect all the FDI points. The user has the option of having the minimum spanning tree formed rectilinearly or using airline distances. Under either case, a single road adjustment factor is applied to convert the ideal routing length into reality.

How Can We Test The Various Approaches?

As has been demonstrated in FCC ex-parte filings as well as internal FCC analysis. the minimum spanning tree has proven to be a valid and valuable measurement of the reality of a model's feeder routing. In fact, a model should estimate a feeder and subfeeder distance that exceeds the MST distance. We have been able to run the MST analysis for the BCPM and the HAI in a number of states. However, we have not had time or the data to test the HCPM.

HAI Analysis

Using the wire centers for a single state, our analysis examined the relationship between the feeder and subfeeder lengths estimated by HAI 5.0a and the MST distance for each of the model's wire centers, *by wire center quadrant*. That is, a MST distance was estimated for the main clusters that fell within, for example, the north quadrant of a wire center where the MST connected the serving areas with each other and the central office. The calculated MST was then compared with the estimated total feeder and subfeeder distance for the quadrant. The results of the analysis are shown in Table 1.

Table 1. Ratio of Feeder & Subfeeder Distance to MST Distance by HAI 5.0a Wire Center Quadrant

	Default	Steering
Maximum	4.56	2.25
Minimum	0.98	0.77
Average	1.47	1.28
Line Weighted Average	1.65	1.33
Coefficient of Variation	25.4 %	15.7 %
Percent of Quadrants for Which Ratio < 1	0.39%	6.0%

An estimated feeder and subfeeder distance to MST ratio of less than 1 characterized 0.39% of the quadrants. And then, only slightly less than 1. Hence, the HAI 5.0a feeder design passes, for the most part, the MST reality check. There are, however, some relatively high ratios with the maximum being 4.6.

Using the HAI default route-to-air ratio of 1.27 and enabling the steering option results in the statistics shown in the last column of Table 1. Although the maximum estimated distance to MST distance ratio is reduced by half, substantially more quadrants do not pass the MST reality test, i.e., 6% versus 0.39%. This suggests that the user should use this option with care and have a good idea what a reasonable, forward-looking route-to-air ratio is for feeder.

This raises an important question. Does the HAI 5.0a perform a check to determine which of the two methods is more efficient? Based on our analysis, we have determined that the default and steering designs are not compared with each other to determine which one yields the more efficient feeder routing. Rather, the user must select a method beforehand for the entire modeled area (company within a state). This does not allow the HAI model to optimize the feeder route for an individual wire center but forces all wire centers in a company to construct feeder in the same way. The BCPM approach is clearly superior in that it optimizes feeder routing for each individual wire center.

BCPM Analysis

The same methodology used in the analysis of HAI 5.0a feeder distance estimation was used to analyze BCPM's feeder distance estimation. The results of this analysis for the same wire centers are shown in Table 2.

Table 2. Ratio of Feeder & Subfeeder Distance to MST Distance
by BCPM Wire Center Quadrant

Maximum	2.25
Minimum	0.90
Average	1.28
Line Weighted Average	1.33
Coefficient of Variation	13.7 %
Number of Quadrants for Which Ratio < 1	1%

An estimated feeder and subfeeder distance to MST ratio of less than 1 characterized 1% of the quadrants. Although the lowest ratio is 0.9, 62.5% of the quadrants with ratios less than 1 have ratios between 0.96 and 1.0. Although BCPM and HAI are relatively equivalent on this MST reality check, it is important to note that the coefficient of variation in the ratio (a measure of dispersion around the average ratio) is much smaller for BCPM than for HAI (default). This indicates that the BCPM results are more consistent than the HAI.

Design of Distribution and Feeder Conclusions and Recommendations

Are either the BCPM or HAI approaches inefficient? We would have expected both models to exceed the minimum spanning tree. As the name indicates, the MST is the minimum distance necessary to connect points. It does not take into consideration rights-of-way, terrain, obstacles, or cost minimization. In fact, if we were to use the rule of thumb that the air to route conversion factor should be close to 1.4, each model seems to be in line with expectations. As tables 1 and 2 indicate, the BCPM 3.1 and HAI 5.0a are similar, on average, in terms of the amount of estimated feeder and subfeeder distance relative to their MST distances. The simple average ratio is 1.3 for BCPM and 1.5 for HAI 5.0a.

What about the HCPM approach? The HCPM proposes a Cost Minimized Spanning Tree (rectilinear and airline) that is then adjusted to account for road routing. Our major concerns with this approach are 1) we have not been able to analyze and compare the routing with the other models on record, and 2) a single road to route adjustment will overcompensate in rural areas and under compensate in urban areas. In theory, we would agree that the minimum spanning road tree (based upon minimization of costs) would produce the most accurate assessment of the possible distance needed to connect points along a right-of way. However, this approach is not in the proceeding. If nothing else, the theoretical approach needs to he modified to account for the differences in the road to route adjustment that can occur in denser and sparser areas.

Reaction to the Current Synthesized Version of the HCPM (version 2.6)

Our critique here is in addition to the Joint Sponsor-provided critique of the actual logic and coding of the modules within the HCPM provided in Sprint's July 31, 1998 Ex Parte filing to Ms. Magalie Roman Salas from Mr. Pete Sywenki (see Attachment B). A review of the FEEDDIST program has already been provided. The review of the CLUSTINTF and VB clustering are near completion.

This critique focuses on the structure and use of the HCPM as a proxy system. In this review, we looked at structure, ease of use, the ability to be maintained, and reliability.

Overall, we believe the HCPM has definite potential. As mentioned earlier in the comments, the HCPM needs intense testing with actual customer location data. In addition, there are logic errors, structure changes, data definitions, data sources, and auditing needs that must be addressed before we can fully endorse HCPM as acceptable for the national universal service model.

Critical Items absent from the current HCPM

1. User interface:
   - flexible program control
   - option to run one or more states
   - option to run one or more companies
   - easy, somewhat error-proof, way for users to edit inputs (e.g. don't have to worry about sort order, etc.)
   - various levels and types of reports
   - ability to have multiple views or scenarios.

2. System layout:
   - directory structure should be developed separating input files from program and other control files.
   - input files should be separated by state

3. System design and methodology
   - code should be written consistently across the three modules, using the same coding standards, naming conventions
   - code should be written in the same programming language to increase interoperability, reduce maintenance costs, make documentation easier, and make audits easier
   - more current programming techniques should be used when possible (e.g. objects)

4. System review/maintenance:
   - there needs to be a way to audit the system and see intermediate results
   - there needs to be a way to balance geocoded data to the actual line counts, by wire center
   - the model needs to be easy for someone to edit the system to create different output files (changing either the data or the format)
   - the reporting module needs to be flexible
   - there needs to be a consistent set of actual customer data, without which, the HCPM, HAI and BCPM alternative approaches cannot be thoroughly compared and analyzed

General comments on HCPM

1. The system lacks structure. All files and programs are located in the same subdirectory, making it difficult to keep track of individual files. There should be a directory structure in place to separate input files and output files by state. The majority of the file handling has to be done manually. Program files and other control files should be located in a separate directory.

2. The system is a conglomeration of various programming languages. This generally results in systems that are very hard to maintain and more subject to error.

3. There is little or no flexibility in controlling the program flow. In HCPM 2.6 a batch file handles system control. In the latest release of the model, the interface assumes that the buttons are pressed sequentially, thus mimicking the original batch file. This is not the most current, more efficient style of programming.

4. The system as it is written cannot be easily audited. As it currently stands, the system is a "black box" with no way to check intermediate results. At the present time, to look at any of the calculations one must have a programming background.

5. There are concerns about model methodology. We have pointed these out in the attached July 31 Ex Parte filing. The Sponsers will submit a critique of the clustering approach in the near future.

6. The system is exceptionally slow. The run times are a function of the large amounts of data to process, but also are a function of the programming style. Data structures are poorly chosen, resulting in excessive looping during processing.

Conclusion

The Joint Sponsors are not at all opposed to the concept of synthesizing a cost proxy model platform utilizing different elements from various models. A synthesis may produce a product better than any of its predecessors. However, the Joint Sponsors do have some concerns about the process. As it concerns geocoding, the Joint Sponsors believe good geocoded data would be beneficial to any proxy model, but the Joint Sponsors have questioned the quality of the currently available geocoded data. As important to the nature of the available data, the Joint Sponsors are even more concerned about accurately locating those customers that are not geocoded. To this end, the Joint Sponsors propose using a "calculated placement" approach as described herein. The Joint Sponsors recommend a geocoding breakpoint (80-85%) below which geocoded data is not used and customer locations are estimated as discussed. Also, it is essential that the location of the points within clusters be maintained, not discarded, as the clusters are used as serving areas in any proxy model. The selected model should optimize the feeder route on an individual wire center basis similar to the BCPM. Even though the HCPM has definite potential, it needs intense testing with actual customer location data. In addition, there are logic errors, structure changes, data definitions, data sources, and auditing needs that must be addressed before the HCPM is acceptable for the national universal service model per the FCC's own criteria.

Exhibit BKS-6

A BCPM Study Using Geocoded Customer Points

As a benchmark of the BCPM model's assignment of customer locations, we ran a test on several wire centers in Florida served by Sprint, using *geocoded* customer points. These customer locations were geocoded from the customer address data in Sprint's own customer files.

The geocoder used was MapMarker 3.3 Plus, a very current GDT-based geocoder. We accepted geocode results "to the doorstep" *and* – where that failed – to the ZIP+4 centroid (the latter usually places a point on the correct side of the street in a single city block). We considered any *other* points *not* to have been geocoded. On this basis, we experienced geocoding success in the 80% to 90% range, depending on wire center.

We used the same BLR wire center boundaries for these wire centers that were used in BCPM, to maintain true comparability. The Sprint customer data is organized on the basis of the *true* wire center to which each customer belongs, not the approximations provided by BLR. Therefore, we used customer data from each true wire center *and* from each wire center around it. After geocoding this customer data, we then determined which of the points fell within each BLR wire center boundary, assigned those points to that wire center, and discarded unassigned geocoded points. In this way, we minimized the potential for any variation in results that might be due to wire center boundary issues instead of pure customer location allocation issues.

In the data collection process, we were able to determine for each residential location the number of subscribing housing units in that structure; thus, we had the information to fill in the "units in structure" table when these locations were later assigned to microgrids. Also, our collection allowed us to determine for each business location its number of business lines; thus, when it came time to assign these locations to microgrids, we had both the business firms and business lines information available.

However, in each wire center studied, the geocoded residential customers were still fewer than the total number of residential customers recorded in BCPM from Census Bureau data; and the total number of business firms were still fewer than the number used in BCPM from data received from PNR. Therefore, in each wire center we were required – for comparability – to generate *surrogate* residential locations (and the number of housing units in each) and business locations (and the number of lines for each) to bring *each* number to the level of those used in BCPM. The percent of locations surrogated was typically in the 10% to 15% range.

The customer data used happily contained at least a 5-digit ZIP code for every customer location. We decided, therefore, to assign surrogate points by ZIP code. For each ZIP code we determined the number of residential locations and units required, and the number of business firms and lines required, to match the number used in BCPM. We then constructed the *road area* for each ZIP code within each wire center by drawing a "buffer" 500 feet on either side of each road segment. We then *randomly* generated, within each wire center / Zip code road area, the appropriate number of residential locations (and housing units) and business locations (and business lines) to match the BCPM numbers.

The remainder of the process was straightforward. In this study, instead of *allocating* Census Bureau and PNR data to microgrids, we simply determined the locations, units, and business lines that were *located* within each microgrid and assigned them to that microgrid. We retained, from the original BCPM runs, the terrain and road information already specified for each microgrid. We then ran the same aggregation and feeder generating programs against those microgrids (now with geocoded customer location content) as were run in preparing the data originally for BCPM. The comma-separated variables files produced became replacement input to the BCPM model.

Exhibit BKS-7

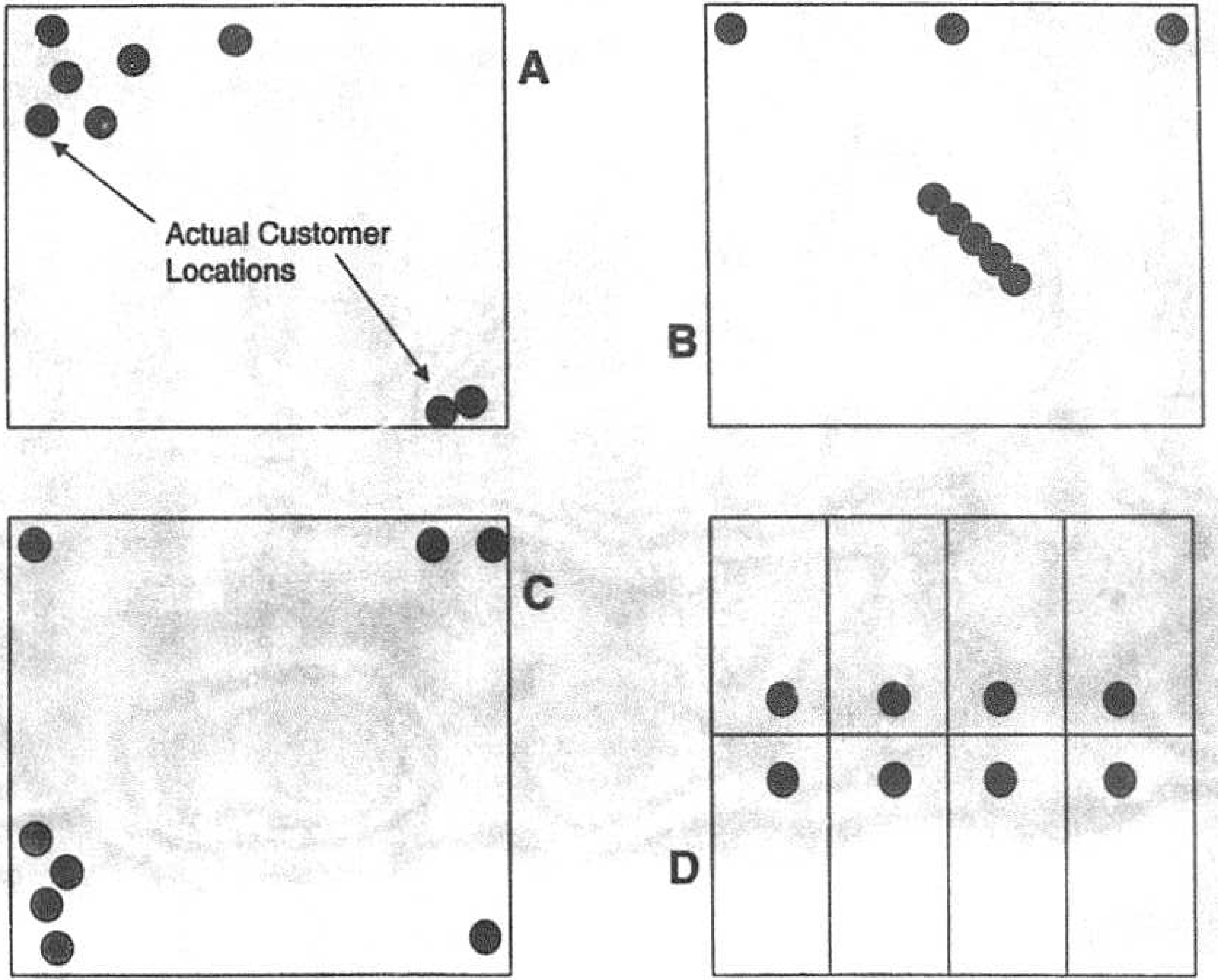
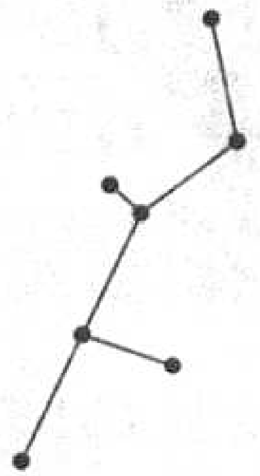

A
Actual Customer Locations
B
C
D

Exhibit BKS-8

Using Minimum Spanning Trees to Estimate Subscriber Dispersion and Minimum Network Length

A *Minimum Spanning Tree* is a construct from graph theory. It is commonly used in network design as a measure of the *dispersion* of the points to be served by a network, and as a benchmark for the *shortest possible* length of a network to serve those points.



A Minimum Spanning Tree

For a set of *points* (we would say "subscriber locations"), a *Spanning Tree* is a set of straight line segments that connect *every* point (subscriber), simply drawing a line from one point to another, using no excess lines. If there are *N* points, there will necessarily be *N - 1* of these line segments.

The *Minimum* Spanning Tree of a set of points is that set of connecting line segments whose total length is the *shortest possible* for this set of points.

If one knows the distance from every point to every other point in a set, it is not difficult to construct, and to determine the length of, the Minimum Spanning Tree of those points. The famous algorithm for calculating it, published in 1957 by R.C. Prim of Bell Labs, uses this simple logic:

- First, find the two points that are closest to each other and connect them
- Then repetitively, until all points have been connected, find the shortest distance between any already-connected point and any not-yet-connected point, and connect those points

As Prim pointed out in his paper, there is one and only one *shortest total length*.

While the Minimum Spanning Tree seems a very satisfying measure of the degree of *dispersion* of a set of points, there are two objections we would make to its use in estimating a minimum possible *telephone network*:

- First, telephone networks are not constructed by chaining together one subscriber to another. Rather, a set of cables is run along as optimal a path as possible, and short drops from *terminals* connect those cables to subscribers. (Those terminals represent additional *points* in the network, introduced at will by the designer.) Perhaps one could construct a *shorter* network than a Minimum Spanning Tree when using this method.

- On the other hand, the line segments of a Minimum Spanning Tree run directly from one point to another. If those points represent real subscribers, these lines could possibly run across back lots and cow pastures, and through lakes, mountains, and tall buildings. Surely the Minimum Spanning Tree is a significant *understatement* of the realistic routing of network cable.

Both points have merit. They are addressed below in order.

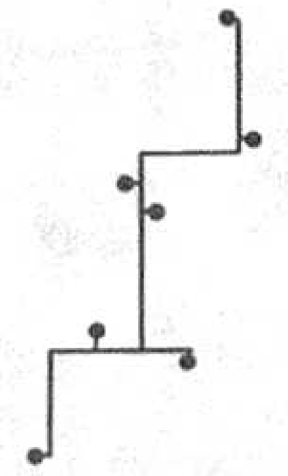
The Minimum Spanning Tree construct does not allow the introduction of additional points. This is what keeps the construct simple, and easy to calculate. The construct that attempts minimum total length by adding additional points as necessary is known as a *Steiner Minimum Tree*, named for the mathematician Jakob Steiner who posed this construction problem in designing road networks two centuries ago.

There are not many configurations of original points for which adding additional points (forming a Steiner Minimum Tree) will connect with less total length than a Minimum Spanning Tree, but there are *some*. Even in those special cases, however, there is an absolute limit to the improvement. In a paper published in 1990, D. Z. Du and Frank Hwang (Hwang is of Bell Labs) proved that adding extra interconnection points cannot reduce the total length of the tree by more than about 13 percent.

The calculation of a Steiner Minimum Tree for a large number of points is known to be a monstrous effort, taking immense amounts of computer time. Because it seldom improves on a Minimum Spanning Tree's length, and even then only slightly, the simple Minimum Spanning Tree calculation is regularly used as a benchmark for *shortest theoretical length.*

The second objection has greater significance, and illustrates why the Minimum Spanning Tree is simply a *benchmark* for, and not a realistic measure of, the shortest possible network. Because a Minimum Spanning Tree has no respect for rights-of-way, and a telephone network *must* respect them, the Minimum Spanning Tree regularly understates the minimum practical network length.

In the figure at the right, we have constructed a more nearly realistic part of a network, running along what would be streets or roads. Even having tailored this sub-network to this exact set of points, we find the length of the tree in this figure to be 18% greater than the length of the Minimum Spanning Tree for those same points. To account for future growth, real telephone networks can not be tailored so tightly to a static set of customers, and are therefore even *less* efficient of length than in the illustration at the right.



A More Realistic Tree

We know that a common rule-of-thumb factor used by telephone engineers to convert arbitrary straight line distances (such as are used in a Minimum Spanning Tree) to realistic cable runs is the square root of 2, or 1.414. It would be no great leap to consider that a reasonable minimum network would be something like 1.414 times the length of the Minimum Spanning Tree of the points served.

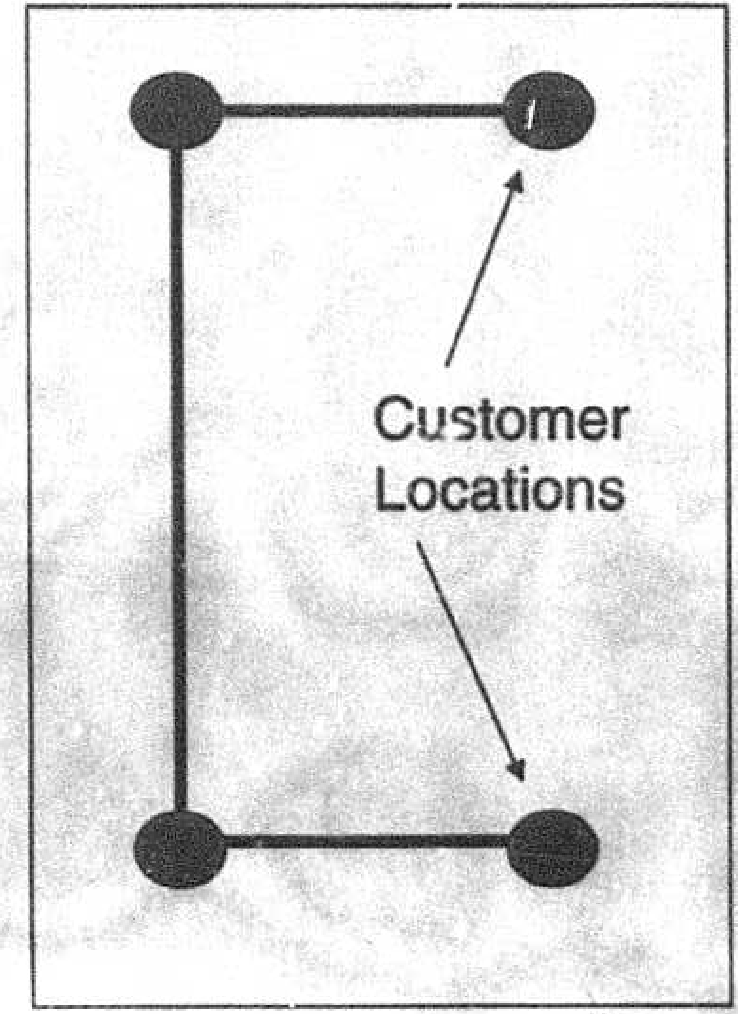
Papers Cited:

R. C. Prim, "Shortest Connection Matrix Network and Some Generalizations," *Bell System Technical Journal*: 36, 1389-1401, November 1957

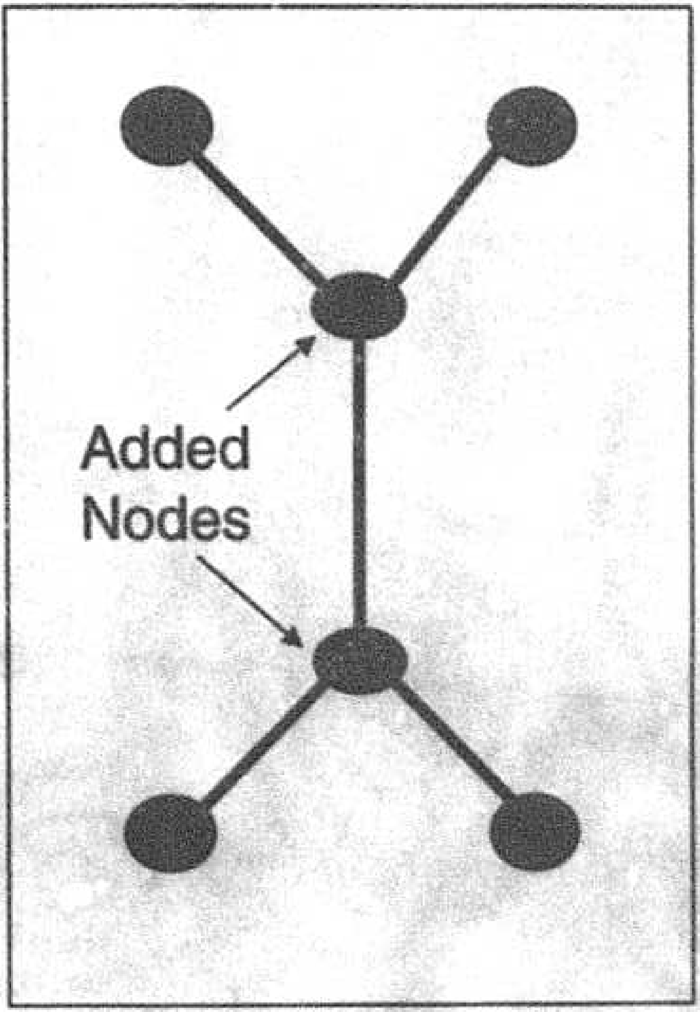
D. Z. Du & Frank Hwang, "A Proof of Gilbert-Pollak's Conjecture on the Steiner Ratio". Publication 90-72 of the *Center for Discrete Mathematics & Theoretical Computer Science of Rutgers University*, 1990

Exhibit BKS-9

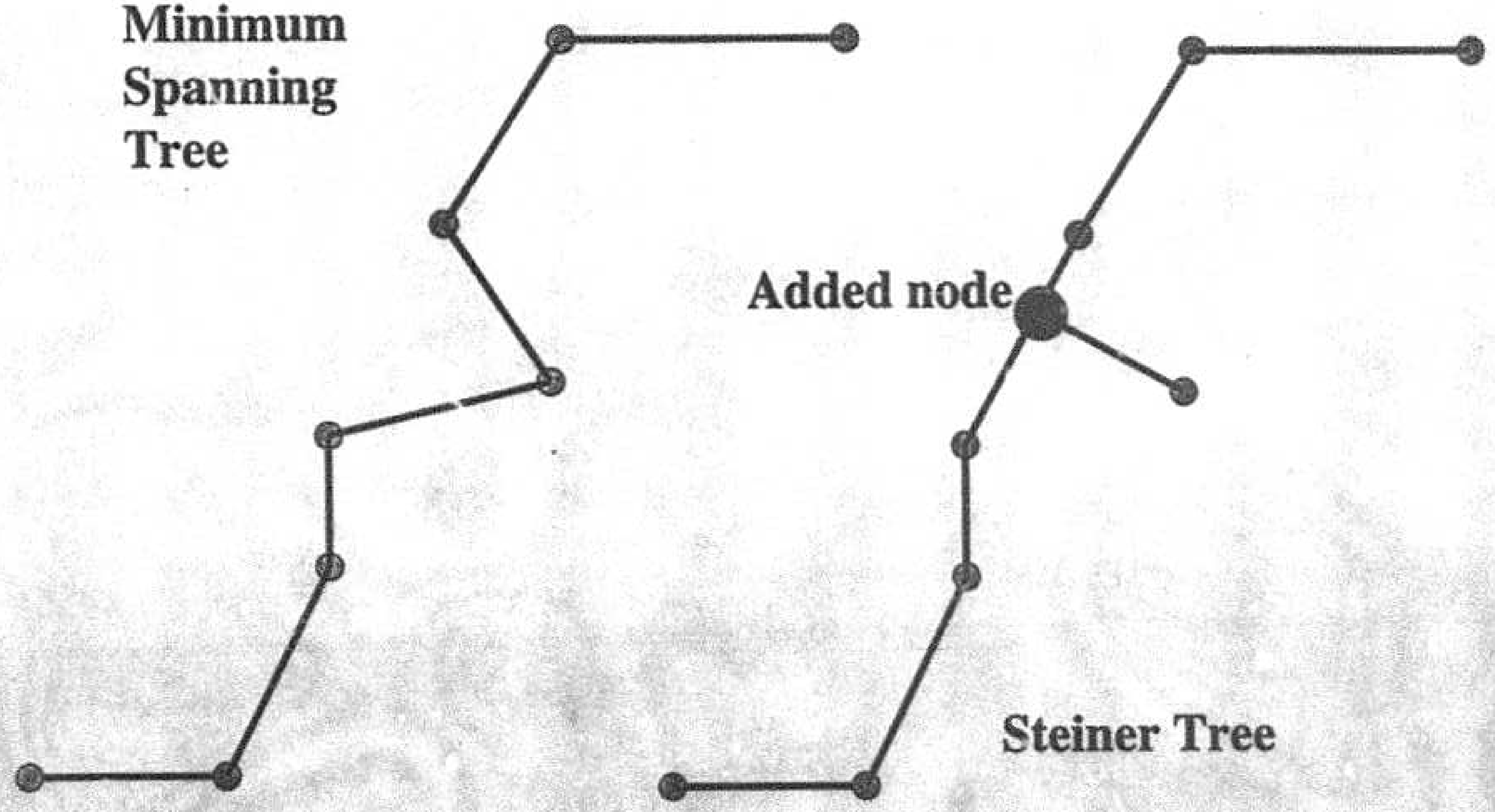
Minimum Spanning Tree



Steiner Tree



The length of a Steiner Tree will always be 87% or more of the length of the associated MST.



Minimum Spanning Tree

Added node

Steiner Tree

The length of a Steiner Tree will always be 87% or more of the length of the associated Minimum Spanning Tree.

Exhibit BKS-10

Minimum Spanning Trees for BCPM: Methodology Used

This paper describes the methodology used to calculate the Minimum Spanning Tree (MST) lengths for BCPM ultimate grids. Because of constraints we explain here, we can not calculate the MST for BCPM in the *classic* manner. But, recognizing that the purpose of this calculation is to determine whether the BCPM Model produces sufficient cable to cover Minimum Spanning Tree length, we employ a methodology that – when it uses statistical estimation – always does so in a direction that may possibly *overstate* the MST length, in order to assure that BCPM is stringently critiqued.

A classic MST calculation operates upon a set of discrete points at specified locations, determines the shortest set of line segments to connect all points, and sums the length of those line segments. In BCPM, there is no set of discrete points. Rather, within each *micrognid* (an area roughly 1500 ft. by 1800 ft., the base statistical unit of the model), there are residential and business locations whose number was determined from underlying Census Block data, apportioned on the basis of road lengths. The position of each location *within* the microgrid is not made specific in the BCPM Model, but it is clear that the customers are assumed to be uniformly spread along the road segments within the microgrid.

Had there been sufficient time to do so, we would have programmatically *placed* the customer locations uniformly along all roads, redrawn the grids, and used *that* set of points and grids for this analysis. But that is a large endeavor, demanding much more time than we were allowed. Therefore, we took a careful mathematical approach that will very closely *estimate* the MST length in a BCPM ultimate grid, always leaning toward the high side. To check our methodology, we did hand placements of points in grids as described here, performed the classic MST calculation on those points, and compared those results to the results of our mathematical approach. In each case, our mathematical approach produced a slightly *longer* Minimum Spanning Tree.

Our approach is essentially this. First, determine the MST and its length *within* each microgrid of the ultimate grid. Then, determine the minimal set of line segments, and their lengths, that can connect the micrograds. Finally, add all these numbers to produce the estimated Minimum Spanning Tree length for the BCPM ultimate grid. (This approach somewhat *overestimates* the MST length because in some cases the point-to-point distance *across* microgrids might be less than the point-to-point distance *within* a microgrid.)

Said another way, we first compute a minimum spanning *subtree* for each occupied microgrid, then connect all the subtrees in an ultimate grid. A BCPM ultimate grid may contain anywhere from 1 to 64 (and, in some cases, a few more) microgrids.

The calculation of the sub-tree for a micro-grid begins by determining the integral number of locations present in that micro-grid. The need for an *integral* number of customer locations is due to the nature of a minimum spanning tree. Specifically, the MST for *n* points must have *n - 1* line segments, or edges, that connect all of those points. If *n*, which is equivalent to our customer locations, is allowed to be fractional, we introduce the possibility of a fractional count of line segments, which does not exist in graph theory.

The basic equation for calculating the number of locations is:

Equation 1.

$$\begin{aligned}\text{Locations} = {} & \text{BusFirms} / 1 + \text{HU1Det} / 1 + \text{HU1Att} / 1 + \text{HU2} / 2 + \text{HU3to4} / 3 + \\ & \text{HU5to9} / 7 + \text{HU10to19} / 15 + \text{HU20to49} / 35 + \text{HU50Plus} / 55 + \\ & \text{HUMbl} / 1 + \text{HUOther} / 1\end{aligned}$$

For the following discussion, the number of locations for an entity (micro-grid or ultimate grid) is the result of equation 1 using the demographic information of that entity. The rounding method used is: round up for numbers with a fractional part that is greater than or equal to 0.5, and round down for numbers with a fractional part less than 0.5.

Micro-grids carry a fractional representation of demographic information, i.e. business firms and housing units. Ultimate grids carry this same demographic information as whole integers, resulting from the rounded sum of demographics from the aggregated micro-grids. One method for determining the integral number of locations might be to simply round the number of locations per each micro-grid. Unfortunately, this can result in a complete loss of locations when compared to the number of locations for the ultimate grid. Consider the case of an ultimate grid with 1 location that is composed of four micro-grids. The number of locations for each micro-grid is 0.1, 0.2, 0.3, and 0.4. Rounding the number of locations for each micro-grid would result in 0 locations for each which, in sum, does not equal the number of locations for the ultimate grid.

To alleviate this problem, we use the following method, which allocates an integral number of locations to the micro-grids that, in sum, equal the number of locations for the ultimate grid. First, calculate the rounded number of locations using the data from the ultimate grid. This is the total number of locations that we can allocate amongst the micro-grids. Next, for each micro-grid, allocate from the total location count based on the rounded number of locations for the micro-grid. This may produce either a shortage of locations, when not enough are allocated, or a surplus, when too many locations are allocated. The following illustrates a four micro-grid case that produces a shortage after the initial allocation.

Ultimate Grid Location Count: 4

Micro-grid	Locations	Allocated
1	1.8	2
2	1.2	1
3	0.4	0
4	0.2	0
Shortage		1

When there is a shortage after the initial allocation, the next step attempts to fairly allocate the remaining locations. This is accomplished by taking all micro-grids whose location count is rounded down and sorting them in descending order by the fractional part of locations. Next, step down the list, allocating one location to each micro-grid until we exhaust the shortage. This is a fair method for allocating the shortage since micro-grids with a more significant sub-0.5 fractional part will be allocated locations from the shortage before those with smaller sub-0.5 fractional parts. Here is the above example after allocating the shortage (where * denotes a candidate for shortage allocation).

Micro-grid	Locations	Allocated
1	1.8	2
2 *	1.2	1
3 *	0.4	1
4 *	0.2	0
Shortage		0

The total number of locations allocated to the micro-grids now equals the number of locations calculated for the ultimate grid. However, it is possible after the initial allocation that a surplus of locations was allocated. The following is an example that results in a surplus.

Ultimate Grid Location Count: 4

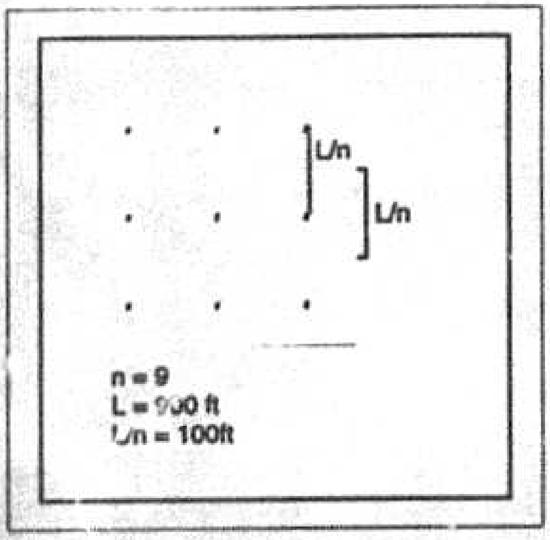
Micro-grid	Locations	Allocated
1	1.6	2
2	1.5	2
3	0.8	1
4	0.4	0
Surplus		-1

To remedy an over-allocation of locations, we use a method that is the converse of that used to allocate any shortage. Specifically, take all of the micro-grids whose location count is rounded up and sort in ascending order by the fractional part of locations. Then step down the list de-allocating 1 location from each micro-grid until the surplus is fully accounted for. Here is the above example after de-allocation (where * denotes a candidate for de-allocation).

Micro-grid	Locations	Allocated
1 *	1.6	2
2 *	1.5	0
3 *	0.8	1
4	0.4	0
Surplus		0

This concludes the method for allocating an integral number of locations to the micro-grids. We fairly account for any allocation shortage or surplus by considering the fractional part of the location count of a micro-grid - a much more fair method than one that randomly allocates or de-allocates locations.

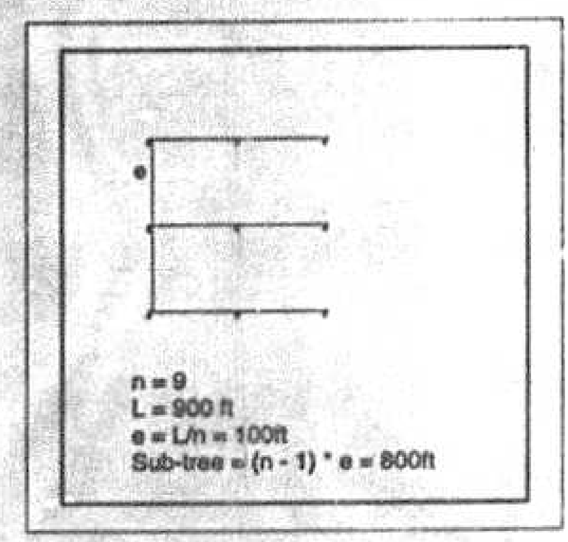
Next, customer locations are mathematically dispersed within the micro-grid. Since BCPM assumes that customers are located along roads, we attempt to logically place customer locations along the roads of a micro-grid. Therefore, to evenly disperse *n* customers along *L* ft of roads, we logically place a customer every *L / n* ft. Or in equivalent terms, we center customers in adjacent customer areas that measure *L / n* ft by *L / n* ft. The *n* square areas are arranged in a group that is centered over the road centroid of the micro-grid. This is easily visualized when the number of locations has an integral square root.



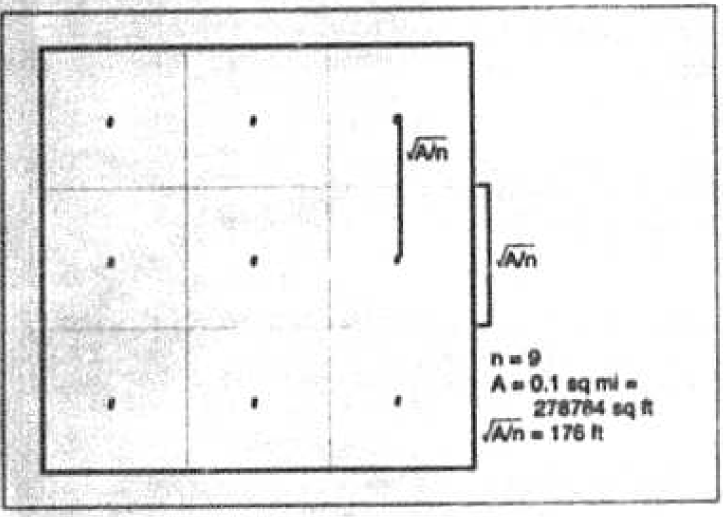
Notice that for any one customer, its nearest neighbor is *L / n* ft away. So the value of L / n not only gives us the dimensions of the customer areas, it also gives us the minimum distance between a customer and its nearest neighbor. This allows us to easily calculate the sub-tree for a micro-grid:

$$\text{Sub-tree} = (n - 1) * e$$

where *e* is equal to *L / n*, *n* is the number of locations, and *L* is the total length of road segments within the micro-grid.



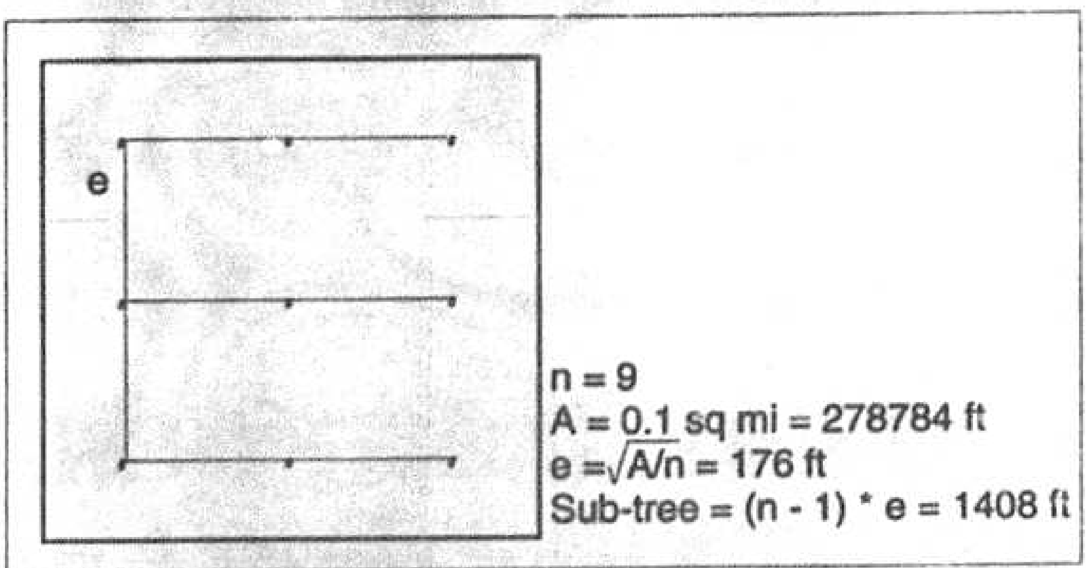
For some micro-grids *L* may be so large that it produces impossibly large customer areas, i.e. areas that in sum have more area than the micro-grid itself. Note the relationship of *e* to *L* for large *L*. For micro-grids a value of L that produces overly large customer areas, we instead, evenly disperse customers using the area of a micro-grid. To mathematically disperse customers by area, we divide the area of micro-grid, *A*, by the number of customer locations, *n*. This gives us *n* customer areas all of which have an area of *A* / *n*. Customers are logically placed in the center of these square areas, and the areas are arranged in a group that is centered over the physical center of the micro-grid (not the road centroid).



As you can see, the distance between a customer and its nearest neighbor is sqrt (A / n), where sqrt () represents the-square-root-of. Thus, when dispersing by area the sub-tree for a micro-grid is:

$$\text{Sub-tree} = (n - 1) * e$$

where *e* is equal to sqrt (*A* / *n*), *n* is the number of locations, and *A* is the area of the micro-grid.



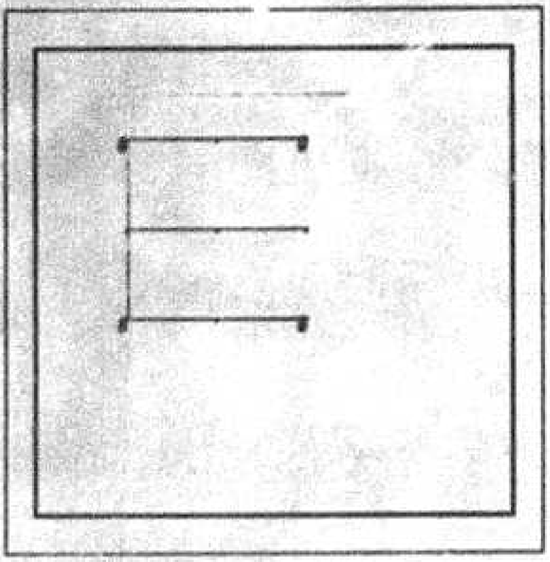
As previously mentioned, dispersion by area is used to realistically limit the size of customer areas. Customers are dispersed by this method only when it produces a customer area size smaller than that produced when dispersing customers by road segment length.

Taking both dispersion methods into account, the mathematical equations for calculating the sub-tree for a micro-grid are:

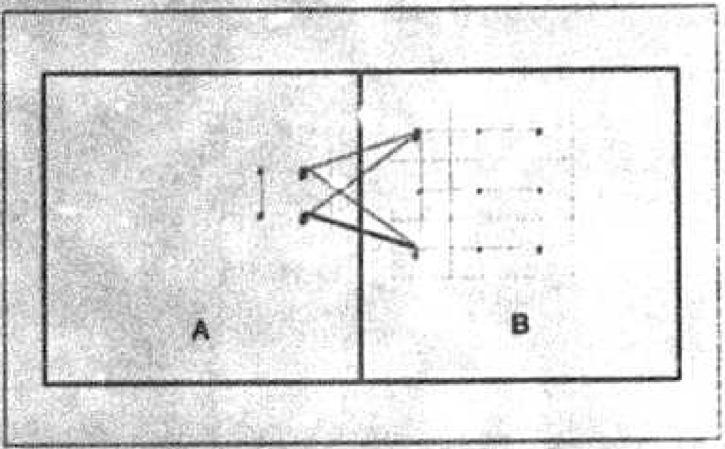
$\text{Sub-tree} = (n - 1) * e$
$e = \min(e^{L}, e^{A})$
$e^{L} = L / n$
$e^{A} = \text{sqrt}(A / n)$

n = customer locations
L = road segment length
A = area

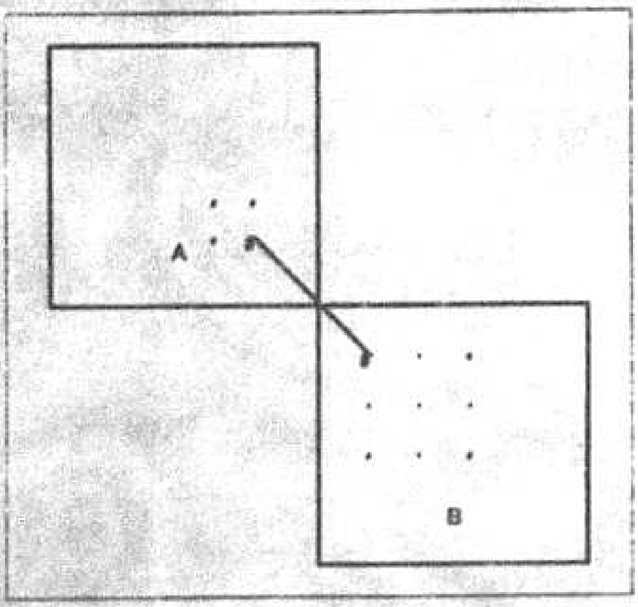
Next, we must interconnect all of the sub-trees. Sub-trees are connected to one another via their corner points.



The corner points that are actually used to connect any two sub-trees depends on the geographic relationship of the two micro-grids in question. If the micro-grids are located directly North/South or East/West of one another, the N/S or E/W corner points are considered for use as the connecting nodes. For example, if micro-grid A is located directly east of micro-grid B, the NW and SW corner points of micro-grid A and the NE and SE corner points of micro-grid B are chosen as candidates for connecting nodes. Of these four candidate nodes, the pair that minimally interconnect the micro-grids are used.



If the two micro-grids do not lie directly N/S or E/W of one another, we use the two corner points of those micro-grid that point toward one another. For example, if micro-grid A lies NW of micro-grid B, the SE corner point of micro-grid A and the NW corner point of micro-grid B are used as connecting nodes.



Using the above inter-distance calculation, we generate the set of edges that interconnect every sub-tree of a micro-grid to every other sub-tree in the ultimate grid. From this set of edges we select *N - 1* edges that minimally interconnect all of the sub-trees, where *N* is the number of sub-trees in the ultimate grid.

When we sum the lengths of all the sub-trees and the lengths of the *N - 1* edges that minimally interconnect the sub-trees, we get the pseudo-minimum spanning tree of our ultimate grid.