Available online at www.sciencedirect.com

SCIENCE DIRECT

Journal of the Franklin Institute 343 (2006) 69–82

Journal of The Franklin Institute

www.elsevier.com/locate/jfranklin

Signal constellations in the hyperbolic plane: A proposal for new communication systems

Eduardo Brandani da Silva[a,*], Marcelo Firer[b], Sueli R. Costa[b], Reginaldo Palazzo Jr.[c]

[a]*Departamento de Matemática, Universidade Estadual de Maringá, Av. Colombo, 5790, 87020-900 Maringá, PR, Brazil*

[b]*Instituto de Matemática, Universidade Estadual de Campinas, Caixa Postal 6065, 13083-970 Campinas, SP, Brazil*

[c]*Faculdade de Engenharia Elétrica e Ciência da Computação, Universidade Estadual de Campinas, Caixa Postal 6065, 13083-970 Campinas, SP, Brazil*

Received 6 September 2005

Abstract

Signal constellations in the hyperbolic plane are provided as an alternative to traditional signal constellations in the Euclidean plane, since channels may actually exist for which the latter signal constellations are not as suitable as the former. A hyperbolic gaussian probability density function, based solely on geometrical considerations, is derived to determine the performance of the hyperbolic signal constellations. Benefits result from an approach conceived in terms of reduced signal-to-noise ratio, needed to achieve a prescribed error rate and equivalent optimum receiver complexity.

Keywords: Modulation; Signal constellations; Hyperbolic space; Hyperbolic gaussian *pdf*

1. Introduction

When designing a new communication system two important aspects to be achieved are: (1) provide coding gain; and (2) be at most as complex as the previously known systems.

*Corresponding author. Tel.: +55 44 30289956.
E-mail addresses: ebsilva@wnet.com.br (E.B. da Silva), mfirer@ime.unicamp.br (M. Firer), sueli@ime.unicamp.br (S.R. Costa), palazzo@dt.fee.unicamp.br (R. Palazzo Jr.).

0016-0032/$30.00
doi:10.1016/j.jfranklin.2005.09.001

In any communication system the information which is transmitted is bounded by a set of impairments from the channel, usually referred to as noise. The characterization of the noise is achieved by knowing its probability distribution function or its probability density function (*pdf*) [1]. As it is well known, this is an important step for obtaining the optimum receiver, and consequently, the undertaking of the communication system. Once the noise has been characterized, the next step is to process properly the signal before transmission so that the noise action on the signal could be efficiently controlled. This step may be done by modulation, coding, and so on. Ref. [2] is a good introduction.

In traditional communication system design the description of the signal constellations is based on vector space techniques with an associated metric. In general, the signal constellation is embedded in a Euclidean space regardless the metric under consideration.

One of objectives in this work is to compare the performance of a communication system whose signal constellation is in the Euclidean and hyperbolic planes. We shall show that the M-PSK and M-QAM signal constellations in the hyperbolic space are an asset when compared to the corresponding signal constellations in the Euclidean space. Partial results were shown in [3–5].

We would like to note that hyperbolic geometry is not obtained from Euclidean geometry and that the former possesses properties which distinguish it from the latter. Two of them are worth mentioning: (i) the Euclide's fifth axiom does not hold, that is, given a straight line and a point external to it, there is more than one straight line, which is parallel to the given straight line; (ii) the sum of the internal angles of a triangle is less than π.

At this point the following question may be asked: Why it is important to consider signal constellations in the hyperbolic plane? One of the main reasons is that, in the hyperbolic space, an infinite number of regular arrangements (tessellations) of points, which are not possible in Euclidean space, may be designed.

Further, a communication system designer has an infinite number of regular hyperbolic tessellations from which he might select the most convenient one. Another reason is that reduced signal-to-noise ratio may be achieved by using signal constellations in the hyperbolic plane when compared with the signal constellations in the Euclidean plane. Furthermore, the signal processing needed in the demodulation process of the signal constellations in the hyperbolic plane is similar to the processing needed in the Euclidean plane.

Moreover, channels may exist for which this geometry is the natural one to work with. For instance, digital data transmission in power transmission lines is an appropriate candidate to use hyperbolic signal constellations, as suggested in [6].

2. Introduction to hyperbolic geometry

In order to present a more self-contained text, a brief summary of hyperbolic geometry will be given in this section. Roughly speaking, hyperbolic geometry is a geometry in which Euclides's fifth axiom (Parallel Postulate), "Given a line and a point external to it, there is an unique line passing through the given point and parallel to the given line", is changed to "Given a line and a point external to it, there are at least two lines passing through the given point and parallel to the given line". In this paper, the axiomatic discussion directly on the models will be avoided and hyperbolic geometry will be introduced analytically. For more details see [7,8].

Some definitions will be now established. Without loss of generality, the two-dimensional Euclidean space, $\mathbf{R}^2$, is considered. However, the definitions may be extended to space $\mathbf{R^n}$, or any Riemman manifold. Let us recall that a squared matrix M is called positive semi-definite if $uMu^t \geqslant 0$ for all vector u.

Let $\Omega \subset \mathbf{R}^2$ be a simply connected subset; thus all simple and closed curves in Ω are contractible to a single point. Now, let $\alpha(t) = (x(t), y(t))$, $t \in (a, b)$ be a parametrizable curve in Ω.

Definition 1. Let $g_{11} = g_{11}(x,y)$, $g_{12} = g_{12}(x,y)$, $g_{21} = g_{21}(x,y)$ and $g_{22} = g_{22}(x,y)$ functions of $\mathbf{R}^2$ in $\mathbf{R}$, such that

$$G(x,y) = \begin{bmatrix} g_{11}(x,y) & g_{12}(x,y) \\ g_{21}(x,y) & g_{22}(x,y) \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$$

is a positive semi-definite matrix for all $(x,y) \in \mathbf{R}^2$. Let $P = (x,y)$ be an arbitrary fixed point in $\mathbf{R}^2$. Given any two vectors u and v in $\mathbf{R}^2$, the inner product of u and v, relative to the point $P = (x,y)$ and the matrix G is defined by

$$\langle u, v\rangle_{G(x,y)} = uG(x,y)v^t.$$

Then,

$$\|v\|_{G(x,y)} = \langle v, v\rangle^{1/2} = [vG(x,y)v^t]^{1/2}$$

is the norm of v in relation to $P = (x,y)$.

Definition 2. Let $\alpha(t) = (x(t), y(t))$, $t \in (a,b)$ be a continuously piecewise differentiable curve in $\mathbf{R}^2$. The "length" $\|\alpha\|$ of α, in relation to the norm given by the matrix $G(x,y)$, is defined by

$$\begin{aligned}
\|\alpha\| &= \int_C \mathrm{d}s = \int_a^b \|\alpha'(t)\|_G \,\mathrm{d}t \\
&= \int_a^b \|x'(t), y'(t)\|_{G(x(t),y(t))} \,\mathrm{d}t \\
&= \int_a^b \sqrt{g_{11}\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\mathrm{d}y}{\mathrm{d}t}(g_{12}+g_{21}) + g_{22}\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2}\,\mathrm{d}t \\
&= \int_a^b \sqrt{g_{11}x'(t)^2 + (g_{12}+g_{21})x'(t)y'(t) + g_{22}y'(t)^2}\,\mathrm{d}t. \qquad (1)
\end{aligned}$$

Thus, given any two points P and Q in $\mathbf{R}^2$, the distance function is defined by

$$\mathrm{d}(P,Q) = \inf_{\gamma} \int_{\gamma} |\gamma'|_G \,\mathrm{d}t,$$

where the infimum takes over all the curves γ connecting P and Q.

It should be observed that the definitions above are locals, that is, they depend on the point P in consideration.

Now, two models of hyperbolic geometry are introduced. Both models are in the complex plane since the hyperbolic isometries have a simpler expression in $\mathbf{C}$.

2.1. Upper-half plane model

The upper-half plane is the set of points $\mathbf{H}^2 = \{\mathbf{x} + \mathbf{iy} : \mathbf{y} > \mathbf{0}\}$. By Definitions 1 and 2 a metric in this set is introduced. In a general way, let $\gamma(t) = (x(t), y(t))$, $a < t < b$ be an arbitrary curve, albeit fixed, as in Definition 1. Considering the matrix

$$G(x,y) = \begin{bmatrix} 1/y^2 & 0 \\ 0 & 1/y^2 \end{bmatrix}, \quad y > 0$$

the length of $\gamma(t)$ is given by

$$\begin{aligned} \|\gamma\| &= \int_a^b \|\gamma'(t)\|_G \, \mathrm{d}t \\ &= \int_a^b \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 \frac{1}{y(t)^2} + \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 \frac{1}{y(t)^2}} \, \mathrm{d}t \\ &= \int_a^b \frac{1}{y(t)} \sqrt{x'(t)^2 + y'(t)^2} \, \mathrm{d}t. \end{aligned}$$

This matrix originates the differential $\mathrm{d}s = |\mathrm{d}z|/\mathrm{Im}[z]$, where $|\mathrm{d}z| = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}$. In this way, given a curve $\gamma : [a,b] \to \mathbf{H}^2$, the length of arc is given by

$$\|\gamma\| = \int_a^b \frac{|\gamma'|}{\mathrm{Im}[\gamma(t)]} \, \mathrm{d}t.$$

Therefore, given z and w in $\mathbf{H}^2$, the distance between them is defined by

$$d(z,w) = \inf_{\gamma} \|\gamma\|, \tag{2}$$

where the infimum takes over all the curves connecting z to w in $\mathbf{H}^2$. Since this expression for the distance between two points is difficult to handle, a more adequate expression may be obtained. Therefore, let us consider the maps $g : \mathbf{H}^2 \to \mathbf{H}^2$ in the form

$$g(z) = \frac{az + b}{cz + d},$$

where a, b, c and d are real numbers and $ad - bc > 0$. An elementary computation yields

$$\frac{|g'(z)|}{\mathrm{Im}[g(z)]} = \frac{1}{\mathrm{Im}[z]}$$

and so

$$\|g(\gamma)\| = \int_a^b \frac{|g'(\gamma(t))| \, |\gamma'(t)|}{\mathrm{Im}[g(\gamma(t))]} \, \mathrm{d}t = \|\gamma\|.$$

Due to this invariance, the invariance of distance d is obtained, namely

$$d(g(z), g(w)) = d(z,w).$$

Thus, the maps g are isometries in the metric space $(\mathbf{H}^2, \mathbf{d})$. Using the above isometries, the following result may be proved (see [7, p. 130]),

Theorem 1. *Let* $z, w \in \mathbf{H}^2$ *and* d *defined as in* (2). *Then*

(i) $d(z, w) = \ln((|z - \overline{w}| + |z - w|)/(|z - \overline{w}| - |z - w|))$;
(ii) $\cosh(d(z, w)) = 1 + |z - w|^2/2\,\mathrm{Im}[z]\,\mathrm{Im}[w]$.

Besides these definitions, the model under analysis is well behaved and an elementary descriptions of all these important features may be found, namely:

1. A geodesic, the shortest path between any two given points, is either an orthogonal line to axis x, with $y > 0$, or an orthogonal semi-circle to axis x.
2. Isometries of $\mathbf{H}^2$ are maps $h(z) = (az + b)/(cz + d)$, where a, b, c and $d \in \mathbf{R}$ are any real numbers such that $ad - bc = 1$.
3. The angle between any two given curves is the same as they are observed by our Euclidean eyes, that is, if $\alpha(t) = a_1(t) + \mathrm{i}a_2(t)$ and $\beta(t) = b_1(t) + \mathrm{i}b_2(t)$ are curves meeting at $x_0 = \alpha(0) = \beta(0)$, then $\cos\theta = \langle \alpha'(0), \beta'(0)\rangle / \|\alpha'(0)\|\,\|\beta'(0)\|$, where θ is the angle between the curves, $\langle\cdot,\cdot\rangle$ is the usual inner product in $\mathbf{R}^2$ and $\|\cdot\|$ the usual norm.

2.2. Unity open disk model

This is Poincaré's unity disk model for the hyperbolic plane. As a point set, it is just the unity open disk $\varDelta = \{z = x + \mathrm{i}y \in \mathbf{C} |\, |\mathbf{z}| = \sqrt{\mathbf{x}^2 + \mathbf{y}^2} < \mathbf{1}\}$. In a similar way to the development of the upper-half plane model, we consider the matrix

$$G(x, y) = \begin{bmatrix} 1/(1 - |z|^2)^2 & 0 \\ 0 & 1/(1 - |z|^2)^2 \end{bmatrix}.$$

This matrix originates the differential

$$\mathrm{d}s = \frac{2|\mathrm{d}z|}{1 - |z|^2},$$

where $|\mathrm{d}z| = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}$.

A metric in $\varDelta$ is defined by

$$d^*(z, w) = \inf \|\gamma\|, \tag{3}$$

where the infimum takes over all curves $\gamma : \mathbf{R} \to \mathbf{\Delta}$ connecting z to w. Then,

Theorem 2. *Let* $z, w \in \varDelta$ *and* d^* *defined as in* (3). *Then*

(i) $d^*(z, w) = \ln((|1 - z\overline{w}| + |z - w|)/(|1 - z\overline{w}| - |z - w|))$;
(ii) $\cosh(d^*(z, w)) = (|1 - z\overline{w}|^2 + |z - w|^2)/(|1 - z\overline{w}|^2 - |z - w|^2)$.

The main facts about the unity open disk model are now enumerated.

1. A geodesic is either the arc of an orthogonal circumference to the unitary circle $\{z \in \mathbf{C} : |\mathbf{z}| = \mathbf{1}\}$ or the diameters of the disk $\varDelta$. In Fig. 1, the arc AB is a hyperbolic line.
2. An isometry of the hyperbolic Poincaré disk is a bijection $T : \varDelta \to \varDelta$ such that $d^*(z, w) = d^*(T(z), T(w))$ for every $z, w \in \varDelta$. There is a simple description of all the

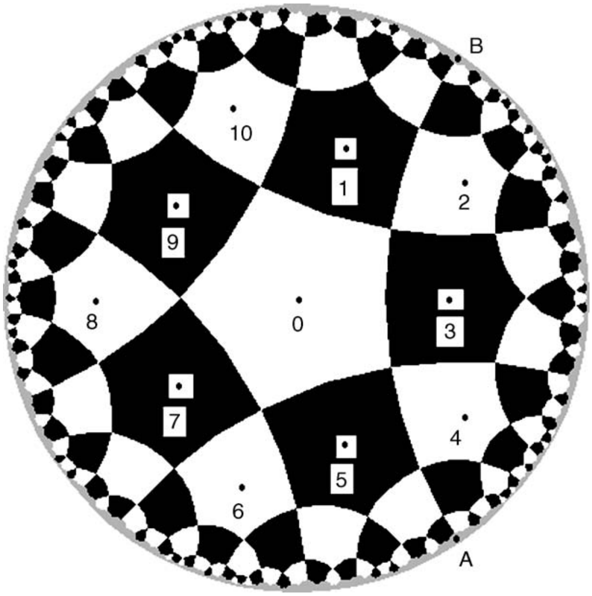

Fig. 1. The tessellation $\{5,4\}$.

(orientation preserving) isometries of Δ: they are just the maps $h(z) = (az + \overline{c})/(cz + \overline{a})$, where $a, c \in \mathbf{C}$ are any complex numbers, such that $|a|^2 - |c|^2 = 1$.

Even though the hyperbolic and the Euclidean spaces have common properties, the existence of more than one parallel line (in fact, infinitely many) is a strong difference between them. Another (equivalent) way to state this difference is to say that the sum of the internal angles of any triangle in Δ is strictly *less* than π.

The map given by $F(z) = (z - i)/(z + i)$ is an isometry between $\mathbf{H}^2$ and Δ. This is an important property, due to the fact that a proper model suitable to a particular situation may be chosen. We denote d^* and d by d_{h}.

Now, if Ω is any region in Δ, then the hyperbolic area of Ω is given by

$$A_{\mathrm{h}} = \iint_{\Omega} \left[\frac{2}{1 - |z|^2}\right]^2 \mathrm{d}x\,\mathrm{d}y.$$

However, if Ω is in $\mathbf{H}^2$, the term $(2/(1 - |z|^2))^2$ is changed by $1/y^2$.

As in the Euclidean case, circles may be defined in Δ. Thus, a circle C with center in w and hyperbolic radius r is defined by

$$C = \{z \in \Delta : d_{\mathrm{h}}(z, w) = r\}.$$

Hyperbolic circles are Euclidean circles too, albeit with different centers and rays.

2.3. Regular tessellations

A *regular tessellation* in the Euclidean and in the hyperbolic planes is a covering of the plane by regular polygons, in which all polygons have the same number of sides.

It may be shown (see [8]) that the number of tessellations in the Euclidean plane is the number of pairs $\{p, q\}$ of integer positive numbers satisfying $(p-2)(q-2) = 4$, where p is the number of sides of each polygon and q is the number of polygons meeting at each vertex. The solutions of the previous equation are $\{3, 6\}$, $\{4, 4\}$ and $\{6, 3\}$. Therefore, the Euclidean plane may be tessellated by equilateral triangles, squares and hexagons. Now, since the sum of the internal angles of any triangle in Δ is strictly less than π, then a $\{p, q\}$ tessellation has to satisfy $(p-2)(q-2) > 4$. Thus, infinite options are available to tessellate the hyperbolic plane. Related to communications systems, this is the main property of the hyperbolic plane, because an infinite number of new signal constellations has to be consider. Fig. 1 shows tessellation $\{5, 4\}$ represented in model Δ. For more details about tessellations, see [7,8].

3. Gaussian probability density function in the hyperbolic plane

One of the contributions of this paper is to derive a gaussian *pdf* in the hyperbolic plane, since the performance analysis of a communication system strongly depends on it. It is usual to consider the noise as a sample of a gaussian process when dealing with the performance analysis of a communication system in Euclidean geometry. This assumption will be kept valid when considering the performance analysis of a communication system in hyperbolic geometry.

The authors in [6] were the first to describe the development of a gaussian measure in the hyperbolic plane starting from a communication model by using waveguides. Subsequently, Terras, in a more detailed work [9], obtained the same hyperbolic gaussian. Researches [6,9] depend in deep results from harmonic analysis in hyperbolic spaces and the hyperbolic gaussian measure obtained by them is very hard to handle in computing systems.

In current research, the derivation of the gaussian *pdf* in the hyperbolic plane will follow a geometric approach rather than a measure theoretical approach.

3.1. Gaussian probability density function in Δ

The gaussian *pdf* in the hyperbolic plane will be derived in this subsection. In the Euclidean case, implicit use of several advantages that $\mathbf{R}^2$ possesses is made due to the fact that it is a normed vector space. In the hyperbolic case, extra attention is needed since it lacks a natural vector space structure as the Euclidean case. Therefore, the geometric properties associated with the gaussian *pdf* in $\mathbf{R}^2$, applicable in the hyperbolic space, are used as often as possible.

The *pdf* of two independent gaussian random variables X and Y, with zero mean and variance σ^2 in the Euclidean space is given by $p(x, y) = (1/2\pi\sigma^2) \exp(-(x^2 + y^2)/2\sigma^2)$.

The term $x^2 + y^2$ may be interpreted as the squared Euclidean distance from (x, y) to the origin. Similarly the function $f(z)$ in Δ is defined as

$$f(z) = B \exp(-A d_{\mathrm{h}}^2(z, 0)), \tag{4}$$

where $z = x + \mathrm{i}y = (x, y)$ is in Δ.

Constants A and B are next determined. Since $f(z)$ is a *pdf*, the volume under the surface generated by $f(z)$ in Δ must be 1, or rather,

$$\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(z)\left(\frac{2}{1-|z|^2}\right)^2 \mathrm{d}y\,\mathrm{d}x = 1.$$

After some algebraic manipulations, we end up with

$$B = \frac{\sqrt{A}}{\sqrt{\pi^3}}\,\mathrm{e}^{-1/4A}\left[erf\left(\frac{1}{2\sqrt{A}}\right)\right]^{-1}. \tag{5}$$

On the other hand, if the gaussian center is not the origin but any other point, say, z_0, then $d_\mathrm{h}^2(z,z_0)$ has to be considered instead of $d_\mathrm{h}^2(z,0)$ in the exponent of (4). We have

$$f(z) = B\exp(-Ad_\mathrm{h}^2(z,z_0)), \tag{6}$$

where B is a constant as in (5).

The gaussian *pdf* in $\mathbf{H}^2$ is easily obtained from (6) by isometry between Δ and $\mathbf{H}^2$. Hence, in $\mathbf{H}^2$, in (6) the term d_h represents the hyperbolic distance (2) between z and z_0.

Eq. (6) describes analytically the gaussian *pdf* in Δ and in $\mathbf{H}^2$. Hence, the appropriate model depends on the problem under consideration, although the computational complexity involved must be taken into account.

Next, two results which characterize geometrically the gaussian *pdfs* in the hyperbolic space will be given.

Theorem 3. *Let $f(z) = B\exp(-Ad_\mathrm{h}^2(z,z_0))$ be the gaussian pdf in $\mathbf{H}^2$. Consider the surface generated by $f(z)$ in $\mathbf{H}^2$. Then the level curves $f(z) = m$ are circles for $m<B$.*

Proof. Let $z_0 = (a,b)$, $z = (x,y)$ points of $\mathbf{H}^2$ and $f(z) = m$, where

$$m = B\exp\left(-A\ln^2\left(\frac{|z-\overline{z_0}|+|z-z_0|}{|z-\overline{z_0}|-|z-z_0|}\right)\right).$$

This implies that

$$\ln^2\left(\frac{|z-\overline{z_0}|+|z-z_0|}{|z-\overline{z_0}|-|z-z_0|}\right) = -\frac{1}{A}\ln\left(\frac{m}{B}\right)>0$$

or rather,

$$\ln\left(\frac{|z-\overline{z_0}|+|z-z_0|}{|z-\overline{z_0}|-|z-z_0|}\right) = \sqrt{-\frac{1}{A}\ln\left(\frac{m}{B}\right)}$$

or equivalently, $d_\mathrm{h}(z,z_0) = r$ where $r = \sqrt{-(1/A)\ln(m/B)}$. Then, $\cosh(d_\mathrm{h}(z,z_0)) = \cosh(r)$. Using equality (ii) of Theorem 1 we obtain

$$1+\frac{|z-z_0|^2}{2\,\mathrm{Im}[z]\,\mathrm{Im}[z_0]} = \cosh(r).$$

Taking $z = (x,y)$ and $z_0 = (a,b)$ we have $|(x-a)+(y-b)\mathrm{i}|/2yb = \cosh(r)-1$. After some algebraic manipulations we obtain

$$(x-a)^2+(y-b\cosh(r))^2 = b^2(\cosh^2(r)-1),$$

which is the equation of a circle in $\mathbf{H}^2$.

Theorem 3 justifies the term hyperbolic gaussian used. □

Theorem 4. *Let $f_1(z) = B\exp(-Ad_h^2(z, z_0))$ and $f_2(z) = B\exp(-Ad_h^2(z, z_1))$ be two gaussian pdfs in $\mathbf{H}^2$ with centers $z_0 = (a, b)$ and $z_1 = (c, d)$, where $z_0, z_1 \in \mathbf{H}^2$, respectively. Let S_1 and S_2 be the surfaces generated by $f_1(z)$ and $f_2(z)$, then the projection in $\mathbf{H}^2$ of the intersection of S_1 with S_2 is an h-line.*

Proof. Let us consider the equality $f_1(z) = f_2(z)$. The equality is valid if, and only if, $d_h^2(z, z_0) = d_h^2(z, z_1)$ and since the distance is a non-negative function we have $\cosh(d(z, z_0)) = \cosh(d(z, z_1))$. Now, by equality (ii) of Theorem 1,

$$1 + \frac{|z - z_0|^2}{2\,\mathrm{Im}[z]\,\mathrm{Im}[z_0]} = 1 + \frac{|z - z_1|^2}{2\,\mathrm{Im}[z]\,\mathrm{Im}[z_1]}. \tag{7}$$

At first, if $\mathrm{Im}[z_0] = \mathrm{Im}[z_1]$, when terms in Eq. (7) are cancelled

$$(x - a)^2 + (y - b)^2 = (x - c)^2 + (y - d)^2,$$

where $b = d$. After some algebraic manipulations, we obtain

$$x = \frac{c + a}{2},$$

which is the equation of the geodesic and, in this case, a Euclidean line. If $\mathrm{Im}[z_0] \neq \mathrm{Im}[z_1]$ in (7), then

$$\frac{(x - a)^2 + (y - b)^2}{b} = \frac{(x - c)^2 + (y - d)^2}{d},$$

or equivalently

$$x^2 - \frac{2x(ad - cb)}{d - b} + y^2 + \frac{(a^2 + b^2)d - (c^2 + d^2)b}{d - b} = 0.$$

This equation represents the equation of a geodesic circle with center on axis x. □

By Theorem 4, the decision thresholds between neighboring signal points of a signal constellation may be determined, and consequently the corresponding Voronoi's regions. Constant A in the *pdf* is determined by the concept of hyperbolic variance, defined below.

3.2. Hyperbolic variance

The concept of hyperbolic variance is used to construct a gaussian *pdf* in the hyperbolic plane with similar properties to a given gaussian *pdf* in the Euclidean plane. Hence, if performance of a communication system using Euclidean signal constellation is compared with a similar one, albeit using a hyperbolic constellation, then we assume that the variance (noise power) is the same for the Euclidean as well as for the hyperbolic gaussian random noise. Theorem 3 shows that the level curves of (6) are hyperbolic circles. It is well known that the level curves of a bivariate Euclidean gaussian *pdf* with equal variances are circles too. Since there is an isometry taking an Euclidean circle to a hyperbolic one, the hyperbolic variance of a gaussian *pdf* $f(z) \equiv f(z, z_0)$ in $\mathbf{H}^2$ is defined by

$$\sigma_h^2 = \iint_{\mathbf{H}^2} \operatorname{arcsinh}(x/y)^2 f(z, i) \frac{\mathrm{d}y\,\mathrm{d}x}{y^2},$$

where $i = (0, 1)$ act like the origin point in $\mathbf{H}^2$ and the term arcsinh in the integration is the hyperbolic distance from $z = (x, y)$ to axis y.

3.3. Channel noise model

The additive white gaussian noise channel model in the Euclidean plane is $y = x + n$, where y is the received signal, x is the transmitted signal and n is a sample of the gaussian random process, may be interpreted as an Euclidean translation (isometry) $T : \mathbf{R}^2 \to \mathbf{R}^2$ given by $y = T(x) = x + n$. In general, $T(.)$ is any Euclidean isometry acting on x.

Since no particular application is being considered, a general context for the channel model should be provided. The previous interpretation of the action of noise on the signal is the most general setting to take into consideration. Therefore, by using the concept of isometry associated with the hyperbolic gaussian noise we have the following.

Definition 3. Let ε be a σ-algebra and Ω the space of events relative to the experiment. A function $Z : (\Omega, \varepsilon) \to \Delta$ is a random variable if, for every subset $A_k = \{\omega \in \Omega : d(Z(\omega), (0, 0)) \leqslant k\}$, $k \in \mathbf{R}$, the condition $A_k \in \varepsilon$ is satisfied.

Hyperbolic Gaussian Noise Hypothesis: Let G_{h} be the set of isometries of $\mathbf{H}^2$. There is a probability space (Ω, ε, P) and a random variable $Z : \Omega \to \Delta$, where $\Omega = \mathbf{H}^2$, ε is the Borel σ-algebra on Ω, and P is a probability measure with $P(g(A)) = P(A)$ for all $A \in \varepsilon$ and $g \in G_h$, such that Z is characterized by the gaussian *pdf* as in (6).

4. Performance analysis

In this section, the performance analysis of M-PSK and M-QAM signal constellations in the hyperbolic plane are analyzed.

4.1. Maximum likelihood receiver

Let $\{s_j\}$, $1 \leqslant j \leqslant L$, be a signal constellation in $\mathbf{H}^2$. Let s_j be the transmitted signal. The received signal is then given by $y = g(s_j)$, where $1 \leqslant j \leqslant L$ and g is a hyperbolic isometry acting as gaussian noise. The hypothesis tests are the following:

$$H_j \longleftrightarrow y = g(s_j), \quad 1 \leqslant j \leqslant L.$$

Geometrically, the receiver decides in favor of signal s_j in the constellation $\{s_j, 1 \leqslant j \leqslant L\}$, if $d_{\mathrm{h}}(y, s_j)$ is the least hyperbolic distance among the signals in the constellation. This is equivalent to take the maximum of the set

$$\left\{ \frac{\mathrm{Re}(\langle y, s_j \rangle)}{\mathrm{Im}(s_j)} - \frac{|y|^2}{2\,\mathrm{Im}(s_j)} - \frac{|s_j|2}{2\,\mathrm{Im}(s_j)} : 1 \leqslant j \leqslant L \right\}. \tag{8}$$

As a consequence (8) provides the mathematical model for the optimum receiver.

4.2. M-PSK signal constellations

In this section, the performance of M-PSK signal constellations in the hyperbolic plane is compared to that in the Euclidean plane. Tables 1 and 2 show that M-PSK signal

Table 1
Performance keeping the same correct probability

Mod.	P_C	E_M^h	E_M^e	d_h^2	d_e^2	G (dB)
4	0.7207	1	2.3116	2.2901	4.2634	0.5880
4	0.9992	9	23.522	28.214	47.044	1.9520
16	0.2481	1	2.5440	0.2066	0.3873	1.3273
16	0.9377	9	90.250	8.1003	13.739	7.7173

Table 2
Correct probabilities for fixed average energy

Modulation	$E_M^e = E_M^h$	P_C^h	P_C^e
4-PSK	1	0.7207	0.5779
4-PSK	9	0.9992	0.9663
16-PSK	1	0.2481	0.1678
16-PSK	9	0.9377	0.4416

constellations in the hyperbolic plane achieve asymptotic gains when compared to M-PSK signal constellations in the Euclidean plane.

The upper-half plane model for the hyperbolic plane is used in this analysis. Let $c_h(i, r_h)$ denote the hyperbolic circle with center $(0, 1)$ and radius r_h. In c_h we choose M equidistant signal points $\{p_1, \ldots, p_M\}$ to form a regular polygon with M sides. These are hyperbolic M-PSK signal constellations. For example, in Fig. 1, the points $1, 3, 5, 7$ and 9 form a 5-PSK constellation in Δ. Similarly in the Euclidean plane, M equidistant signal points $\{q_1, \ldots, q_M\}$ in a circle with radius r_e are considered. Let $p_i = (a_i, b_i)$ and $q_i = (c_i, d_i)$ be the coordinates of the corresponding signal points.

Let $p_i(q_i)$ be the signal point to be transmitted. The correct probability of receiving $p_i(q_i)$ has to be determined and we assume the receiver uses the maximum likelihood decision criterion. From the standard hypothesis testing procedures, the Voronoi region of each signal point $p_i(q_i)$ is obtained. The correct probability associated with the signal constellations in the hyperbolic and Euclidean planes are determined by using standard integration techniques. The corresponding density functions used in the calculation of each one of the correct probabilities are

$$f_{ei}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x - c_i)^2 - \frac{1}{2}(y - d_i)^2\right)$$

in the Euclidean plane, where $q_i = (c_i, d_i)$, and

$$f_{hi}(z) = 0.184164 \exp(-0.73054 d_h^2(z, z_i))$$

in the hyperbolic plane, where $z = (x, y) \in \mathbf{H}^2$. Constants A and B in $f_{hi}(z)$ are determined by assuming the hyperbolic variance of the hyperbolic gaussian *pdf* to be 1.

Let us consider the hyperbolic 4-PSK constellation, given by signals z_0, z_1, z_2 and z_3 arranged counterclockwise on the circumference and assume that one of these signals is to be transmitted over a communication channel whose noise is gaussian. According to the assumption that the hyperbolic noise in $\mathbf{H}^2$ is an isometry, the corresponding hypothesis tests are

$$H_i: \quad y = g(z_i), \quad 1 \leqslant i \leqslant 4,$$

where z_i is the signal point of 4-PSK constellation and $g(.)$ is an isometry that acts as a sample of hyperbolic gaussian noise.

Assuming that the maximum likelihood demodulation process is used, the hypothesis test is given by

$$H_i \longleftrightarrow f_{hi}(z) = 0.184164 \exp(-0.73054 d_h^2(z, z_i)).$$

Let us consider the decision with respect to z_0. The neighbors of z_0 are z_1 and z_3. Let us start with z_1. A decision in favor of H_0 is made if $f_1(y)/f_0(y) \leqslant \lambda \longrightarrow -\frac{1}{2}d_h^2(z, z_1) + \frac{1}{2}d_h^2(z, z_0) \leqslant \ln \lambda$.

If the signals are equally likely, we have $\lambda = P(H_0)/P(H_1) = \frac{1}{4}/\frac{1}{4} = 1$. Thus, $\ln \lambda = 0$, and consequently, the decision threshold is 1. Therefore, $d_h(z, z_0) \leqslant d_h(z, z_1)$. The points $z \in \mathbf{H}^2$ satisfying $d_h(z, z_1) = d_h(z, z_0)$ are exactly the ones which constitute the bisector line of the segment joining z_0 to z_1. Hence, inequality $d_h(z, z_0) \leqslant d_h(z, z_1)$ determines the set of points in the interior of the hyperbolic line (geodesic) of the bisector of z_0 and z_1, that is, the region containing z_0. Note that this procedure is similar to the one employed in the construction of the Voronoi regions in the Euclidean plane.

In this way, the optimum receiver collects a sample from the received signal $y(t)$, $y(t = t_k) = y_k$, in the channel output, and verifies in which region y_k is, and decides by hypothesis H_j.

To determine P_c^h the volume of each hyperbolic gaussian *pdf* has to be found with each signal in the constellation as center and the geodesics of each Voronoi region as boundaries. Since the signals are equiprobable, it is sufficient to find the volume of one region. Hence, $P_c^h = \iint_R f_h(x, y)(1/y^2)\,dy\,dx$, and from extensive calculations, we may see that M-PSK signal constellations in $\mathbf{H}^2$ achieve gains when compared to M-PSK signal constellations in the Euclidean plane.

Table 1 shows the gains associated with each M-PSK signal constellations. For a fixed correct probability, $P_C = P_C^h = P_C^e$, the average energy needed and the squared minimum distance between signal points in the corresponding signal constellations are shown. The figure of merit, G, is given by $G = 10 \log[\frac{d_h^2}{E_M^h} \frac{E_M^e}{d_e^2}]$, where d_h^2 and d_e^2 denote the squared minimum distance, and E_M^h and E_M^e denote the average energy in the hyperbolic and Euclidean planes, respectively. As may be noted, the asymptotic gains achieved by use of signal constellations in the hyperbolic plane are 2.3 dB for the 4-PSK, 12 dB for the 16-PSK, and higher than 16 dB for the 64-PSK, when compared to the corresponding ones in the Euclidean plane.

Table 2 shows the correct probabilities associated with the signal constellations when average signal energy is fixed. As may be noted, the signal constellations in the hyperbolic plane improve considerably the performance of the communication system.

4.3. QAM-like signal constellations

In the Euclidean plane the M-QAM constellations are finite sets of the lattice $\mathbf{Z}^2$ (the tessellation $\{4, 4\}$). Since in the hyperbolic plane this lattice cannot be reproduced, new signal constellations similar to the Euclidean one must be constructed. For a specified number of signal points in $\mathbf{Z}^2$, a subset must be chosen with the same number of signal points in a hyperbolic tessellation. For example, in Fig. 1, points $1, \dots, 11$ may be taken like a 11-QAM in the hyperbolic plane.

In Δ, the tessellations $\{p, 3\}$ are considered in which $p > 6$. For a fixed value of p, let P_0 be a p-sided polygon, called the *fundamental polygon*. The first level of $\{p, 3\}$ consists of P_0 and all the remaining polygons have some intersection with P_0. Let us denote this set by L_1. In this first level, let us consider the set whose elements are the centers of the polygons in set L_1. This set is the QAM-like signal constellation in Δ and the number of elements in this set is $p+1$. Then for a given value of n, $n > 6$, it is always possible to find a hyperbolic constellation with n signal points. Since the performance of 16 and 64-QAM are compared, the first level of the hyperbolic tessellations $\{15, 3\}$ and $\{63, 3\}$ ($p = n-1$) are hyperbolic QAM-like signal constellation.

In order to calculate the correct probabilities, the fundamental polygon of tessellation $\{15, 3\}$ has its center at the origin and 15 signal points are in a hyperbolic circle with radius r_{h}. The noise in the Euclidean plane is gaussian with zero mean and variance 1, and the *pdf* associated with 16 and 64-QAM constellations are given by

$$g_j(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x - x_i)^2 - \frac{1}{2}(y - y_j)^2\right),$$

where $P_j = (x_j, y_j)$ are the coordinates of the signal points. The noise in the hyperbolic plane is also gaussian, albeit with hyperbolic variance 1. The hyperbolic gaussian *pdf* with center in each one of the signal points in Δ is given by

$$f_j(z) = 0.184164 \exp(-0.73054 d_{\mathrm{h}}^2(z, w_j)),$$

where $z = (x, y)$, and $w_j = (a_j, b_j)$, $0 \leqslant j \leqslant m$ where $m = 15, 63$, are the coordinates of the signal points. The constants A and B in $f_j(z)$ are determined in such a way that the hyperbolic variance of the hyperbolic gaussian *pdf* to be 1. The probabilities are calculated by $P_{\mathrm{c}} = \iint_{R_j} f_j(z)[2/(1 - |z|^2)]^2\, \mathrm{d}z$, where R_j is the Voronoi region associated with w_j.

The average energy of an M-QAM constellation, that is, $\{p_1, \dots, p_M\}$, either Euclidean or hyperbolic, is given by $EM = (1/M)\sum_{j=1}^{M} d^2(p_j, 0)$, where d is the distance function between any two given points.

Table 3 illustrates the results obtained from the comparison of the signal constellations in the Euclidean and hyperbolic planes, in which the figure of merit is $G = 10\log[\frac{d_{\mathrm{h}}^2}{E_{\mathrm{M}}^{\mathrm{h}}} \frac{E_{\mathrm{M}}^{\mathrm{e}}}{d_{\mathrm{e}}^2}]$ and "No." is the number of points in the QAM constellations.

Table 4 illustrates the correct probabilities in the hyperbolic and in the Euclidean planes for a fixed value of the average energy.

Table 3
Gain with the same correct probability

No.	$P_{\mathrm{C}}^{\mathrm{h}} = P_{\mathrm{C}}^{\mathrm{e}}$	$E_{\mathrm{M}}^{\mathrm{h}}$	$E_{\mathrm{M}}^{\mathrm{e}}$	d_{h}	d_{e}	G (dB)
16	0.8179	8.7131	22.80	1.007	3.02	5.36
64	0.9997	35.3507	457.38	1.093	6.6	4.49

Table 4
Performance analysis in the hyperbolic and in the Euclidean planes

No. of points	$P_{\mathrm{C}}^{\mathrm{h}}$	$P_{\mathrm{C}}^{\mathrm{e}}$	$E_{\mathrm{M}}^{\mathrm{h}} = r_{\mathrm{h}}^2 = E_{\mathrm{M}}^{\mathrm{e}} = r_{\mathrm{e}}^2$
16-QAM	0.81799	0.54325	8.71317
64-QAM	0.99975	0.47050	35.3507

More complicated finite sets of points in other hyperbolic tessellations may be considered as signal constellations. In this case, the first question is to control the number of points in the sets. A partial answer to this problem is given in [10].

5. Conclusions

In this paper, we have proposed the use of signal constellations in the hyperbolic plane as a means to achieve better performance when transmitting digital signals, for instance, in power transmission lines.

Acknowledgment

The authors wish to thank the referees for pertinent comments and suggestions.

References

[1] A. Papoulis, Probability, Random Variables and Stochastic Process, McGraw-Hill, New York, 1965.
[2] J.G. Proakis, Digital Communications, second ed., McGraw-Hill, New York, 1989.
[3] E.B. Silva, R. Palazzo Jr., S.R. Costa, Improving the performance of asymmetric M-PAM signal constellations in Euclidean space by embedding then in hyperbolic space, Proceedings 1998 IEEE Information Theory Workshop, Killarney, Ireland, June 22–26, 1998, pp. 98–99.
[4] E.B. Silva, R. Palazzo Jr., M-PSK signal constellations in hyperbolic space achieving better performance than the M-PSK signal constellations in Euclidean space, 1999 IEEE Information Theory Workshop, Metsovo, Greece, June 27–July 1, 1999.
[5] E.B. Silva, R. Palazzo Jr., M. Firer, Performance analysis of QAM-like constellations in hyperbolic space, 2000 International Symposium on Information Theory and its Applications, Honolulu, USA, November 5–8, 2000, pp. 568–571.
[6] M.E. Gertsenshtein, V.B. Vasil'ev, Waveguides with random inhomogeneties and Brownian motion in the Lobachevsky plane, Theory Probab. Appl. 4 (1959) 391–398.
[7] A.F. Beardon, The Geometry of Discrete Groups, Springer, Berlin, 1982.
[8] P.A. Firby, C.F. Gardiner, Surface Topology, second ed., Ellis Horwood, Chichester, UK, 1991.
[9] A. Terras, Harmonic Analysis on Symmetric Spaces and Applications I, Springer, Berlin, 1985.
[10] E.B. Silva, R. Palazzo Jr., M. Firer, Counting domains in $\{p, q\}$ tessellations, Appl. Math. Lett. 16 (2003) 323–328.