

Égalisation et synchronisation utilisant un tatouage comme signal de référence

Mário ULIANI NETO¹, Leandro DE C. T. GOMES¹, João Marcos T. ROMANO², Madeleine BONNET³

¹Centre de Recherche en Télécommunications (CPqD)

Rod. Campinas–Mogi-Mirim (SP 340), km 118,5, 13086-902 Campinas, SP, Brésil - tél. +55 19 3705-6724

²Laboratoire de Traitement du Signal Appliqué aux Communications (DSPCom)
Université de Campinas (Unicamp), C.P. 6101, 13083-970 Campinas, SP, Brésil

³UFR de Mathématiques et Informatique - Université René Descartes (Paris V)
75270 Paris Cedex 06, France

uliani@cpqd.com.br, tgomes@cpqd.com.br

romano@decom.fee.unicamp.br, bonnet@math-info.univ-paris5.fr

Résumé – Nous présentons une méthode d'égalisation adaptative et synchronisation basée sur l'emploi d'un tatouage comme signal de référence. Contrairement aux méthodes adaptatives traditionnelles, où l'émission d'information utile est interrompue périodiquement pour transmettre une séquence d'apprentissage, le tatouage est émis continuellement à travers le canal de communication et atteint le récepteur en même temps que le signal portant l'information utile. Si l'on connaît le tatouage original au récepteur, il est possible d'adapter continuellement les coefficients du filtre égaliseur par l'analyse du signal de tatouage reçu de façon à compenser l'action du canal sur le signal émis.

Abstract – We present a method for adaptive equalisation and synchronisation based on the use of a watermark as reference signal. Unlike traditional adaptive methods, in which transmission of useful information is interrupted periodically to transmit training sequences, the watermark is continuously transmitted through the communication channel and reaches the receiver at the same time as the signal that carries useful information. If the original watermark is known to the receiver, the coefficients of the equalisation filter may be continuously adapted by analysing the received watermark in order to reverse channel effects on the transmitted signal.

1 Introduction

Les méthodes traditionnelles d'égalisation adaptative utilisent des séquences d'apprentissage comme signal de référence. Dans ces méthodes, l'émission d'information utile est interrompue périodiquement pour transmettre la séquence d'apprentissage. La comparaison entre la séquence reçue et la séquence originale (connue du récepteur) permet d'adapter les coefficients du filtre égaliseur.

Pour certaines applications, l'interruption périodique de l'émission d'information utile peut être indésirable ou même inacceptable. Dans ces cas, des méthodes d'égalisation aveugle peuvent être utilisées, quoique souvent avec une complexité de calcul plus élevée et une performance moins satisfaisante.

Notre méthode d'égalisation est basée sur l'utilisation d'un *tatouage* comme signal de référence [1, 2]. Le tatouage est un signal de nature similaire à celle du signal portant l'information utile et que l'on insère dans ce dernier. Les deux signaux atteignent simultanément le récepteur. Le tatouage original étant connu au récepteur, il est possible d'adapter continuellement les coefficients du filtre égaliseur par l'analyse du signal de tatouage reçu. Contrairement aux méthodes d'égalisation traditionnelles, le signal portant l'information utile n'est jamais interrompu ; en revanche, une partie de la puissance du signal émis doit être

consacrée en permanence au signal de tatouage.

Le processus de détection du tatouage nécessite une synchronisation précise. Nous proposons une technique de synchronisation utilisant un algorithme de programmation dynamique et une séquence d'apprentissage contenue dans le tatouage.

Un avantage additionnel de notre approche est la compatibilité avec des systèmes existants qui ne sont pas conçus pour détecter le tatouage. Dans ce cas, le tatouage est vu comme du bruit et doit être assez faible pour ne pas avoir d'influence significative sur la performance du détecteur. Cela est appelé la propriété de *transparence* du tatouage. Cette propriété peut être satisfaite en rendant le rapport signal à tatouage (RST, rapport entre la puissance du signal portant l'information utile et celle du tatouage) suffisamment élevé. Lorsqu'il s'agit de signaux qui présentent une interprétation sensorielle (e.g. audio, images, vidéo), des modèles perceptifs peuvent être utilisés pour assurer la transparence du tatouage.

Une approche similaire à la nôtre a été proposée sous le nom d'*apprentissage superposé* (*superimposed training* en anglais) [3]. Cette approche consiste aussi à superposer un signal de référence au signal portant l'information utile ; néanmoins, ce signal n'est en général pas transparent, ce qui rend le système incompatible avec des récepteurs qui ne sont pas adaptés à la technique.

Dans l'exposé du principe d'égalisation avec tatouage, nous présentons le cas de signaux non corrélés puis corrélés. Notre méthode de détection et synchronisation du tatouage est alors exposée. Nous terminons par des résultats de simulation comportant des pertes d'échantillons et des perspectives de poursuite de ces recherches.

2 Égalisation adaptative utilisant un tatouage

La figure 1 illustre le schéma d'égalisation proposé. Le signal de tatouage $m(n)$ est une séquence pseudo-aléatoire blanche de variance unitaire. Il est superposé au signal $t(n)$ portant l'information utile : $s(n) = t(n) + \alpha_m m(n)$, où $s(n)$ est le signal émis et α_m un facteur d'échelle. Le signal $s(n)$ est transmis continuellement à travers un système que l'on suppose linéaire et à réponse impulsionnelle finie (FIR) avec fonction de transfert d'ordre N . Du côté du récepteur, le tatouage est utilisé comme signal de référence pour l'égaliseur, ce qui amène à la fonction coût $J = E[|y(n) - \alpha_m m(n)|^2]$ (critère de Wiener), où $y(n)$ est le signal à la sortie de l'égaliseur. En minimisant cette fonction par rapport aux coefficients \mathbf{w} du filtre égaliseur, on obtient les coefficients optimaux \mathbf{w}_o .

La matrice de corrélation \mathbf{R} de dimension $M \times M$ à l'entrée de l'égaliseur est exprimée par :

$$\mathbf{R} = \alpha_m^2 \mathbf{H} \mathbf{E}[\mathbf{m}(n) \mathbf{m}^T(n)] \mathbf{H}^T + \mathbf{H} \mathbf{E}[\mathbf{t}(n) \mathbf{t}^T(n)] \mathbf{H}^T \quad (1)$$

où α_m^2 est la variance du tatouage, $\mathbf{H} \in \mathbb{R}^{(M+N+1) \times (M+1)}$ la matrice de convolution du système, et $\mathbf{m}(n)$ et $\mathbf{t}(n)$ des vecteurs de dimension $1 \times M$ contenant M échantillons des signaux $m(n)$ et $t(n)$. Le vecteur de corrélation croisée \mathbf{p} à l'égaliseur est défini par :

$$\mathbf{p} = \alpha_m^2 \mathbf{H} \mathbf{E}[\mathbf{m}(n) m(n-d)] \quad (2)$$

où d est le délai optimal d'égalisation (fourni par l'algorithme de synchronisation). Les coefficients optimaux seront alors :

$$\mathbf{w}_o = \mathbf{R}^{-1} \mathbf{p} = \alpha_m^2 (\alpha_m^2 \mathbf{H} \mathbf{E}[\mathbf{m}(n) \mathbf{m}^T(n)] \mathbf{H}^T + \mathbf{H} \mathbf{E}[\mathbf{t}(n) \mathbf{t}^T(n)] \mathbf{H}^T)^{-1} \mathbf{H} \mathbf{E}[\mathbf{m}(n) m(n-d)] \quad (3)$$

Lorsque le tatouage $m(n)$ est un signal blanc et d'énergie unitaire, le terme $E[\mathbf{m}(n) \mathbf{m}^T(n)]$ devient une matrice identité. Si l'on suppose que $t(n)$ est aussi non corrélé (e.g. un signal numérique), le terme $E[\mathbf{t}(n) \mathbf{t}^T(n)]$ devient une matrice identité multipliée par la variance σ_t^2 du signal $\mathbf{t}(n)$. La matrice \mathbf{R} peut ainsi être écrite comme :

$$\mathbf{R} = \alpha_m^2 \mathbf{H} \mathbf{H}^T + \sigma_t^2 \mathbf{H} \mathbf{H}^T = (\alpha_m^2 + \sigma_t^2) \mathbf{H} \mathbf{H}^T \quad (4)$$

Alors, la matrice \mathbf{R} ne dépend plus du signal $\mathbf{t}(n)$ (à l'exception de sa variance σ_t^2 , qui peut être estimée). Les coefficients optimaux \mathbf{w}_o et le signal d'erreur dans la condition optimale $e_0(n)$ peuvent être exprimés par :

$$\mathbf{w}_o = \frac{\alpha_m^2}{\alpha_m^2 + \sigma_t^2} (\mathbf{H} \mathbf{H}^T)^{-1} \mathbf{H} \mathbf{E}[\mathbf{m}(n) m(n-d)] \quad (5)$$

Les coefficients de l'égaliseur tendent donc vers un point qui inverse la réponse impulsionnelle du système à un facteur d'échelle près.

On considère maintenant le cas où le signal $t(n)$ est corrélé (e.g. un signal audio). Nous avons alors :

$$\mathbf{R} = \alpha_m^2 \mathbf{H} \mathbf{H}^T + \mathbf{H} \mathbf{E}[\mathbf{t}(n) \mathbf{t}^T(n)] \mathbf{H}^T \quad (6)$$

Pour que l'égaliseur converge vers la solution de Wiener, une méthode de blanchiment doit être utilisée. On utilise un filtre d'erreur de prédiction à la sortie de l'égaliseur (figure 1) dont le but est de blanchir le signal $y(n)$ à l'aide du filtrage de ses composantes périodiques. La sortie du prédicteur correspondra à la partie corrélée du signal $y(n)$. La différence entre $y(n)$ et le signal à la sortie du prédicteur fournit $y'(n)$, correspondant aux composantes non corrélées à la sortie de l'égaliseur :

$$y'(n) = y(n) - \sum_{k=1}^P w_{f,k} y(n-k-\Delta) \quad (7)$$

où $w_{f,k}$ sont les coefficients du prédicteur avec P éléments de retard. Le signal d'erreur $y'(n)$ tend alors vers les composantes non corrélées de $t(n)$, plus le tatouage $m(n)$ (déjà non corrélé). Le délai Δ doit être suffisant pour permettre d'enlever la corrélation du signal $y(n)$.

3 Détection et synchronisation du tatouage

Nous décrivons ci-dessous le processus de construction et détection du signal de tatouage, ce qui comprend un algorithme permettant au système d'effectuer la synchronisation initiale, c'est-à-dire trouver le début des symboles composant le tatouage, et de resynchroniser lorsque la phase du signal change (e.g. à cause d'un retard introduit par des problèmes de transmission).

Un dictionnaire $\hat{\mathbf{C}}$ est utilisé pour construire le signal de tatouage $m(n)$. Il contient $\hat{\mathbf{u}}_k = [\hat{u}_k(0) \cdots \hat{u}_k(N-1)]$ ($k \in [0, \hat{K}-1]$) vecteurs associés à \hat{K} symboles [4, 5]. La séquence de symboles $\mathbf{z} = [z_0 \cdots z_{M-1}]$ qui composent le tatouage est obtenue selon la règle $z_j = j \bmod \hat{K}$, où z_j est le j -ème symbole dans le signal de tatouage ($j \in [0, J-1]$).

À la détection, une fenêtre glissante est utilisée pour calculer N mesures de corrélation pour chacun des J symboles dans le tatouage $m(n)$ et pour chacun des \hat{K} vecteurs dans le dictionnaire :

$$\dot{r}(\lambda, k, j) = \left| \sum_{n=0}^{N-1} x(jN+n+\lambda) \hat{u}_k(n) \right| \quad (8)$$

où $\lambda \in [-\Lambda, \Lambda-1]$ est le décalage de la fenêtre glissante ($\Lambda = N/2$ pour N pair) et $x(n)$ est le signal reçu. En maximisant cette fonction par rapport à k , on construit deux matrices $\mathbf{A} = \{\beta_{\lambda,j}\}$ et $\mathbf{B} = \{\chi_{\lambda,j}\}$ dont les lignes correspondent au décalage λ et les colonnes à la position j du symbole dans le signal de tatouage :

$$\beta_{\lambda,j} = \max_k \dot{r}(\lambda, k, j) \quad \text{et} \quad \chi_{\lambda,j} = \arg \max_k \dot{r}(\lambda, k, j) \quad (9)$$

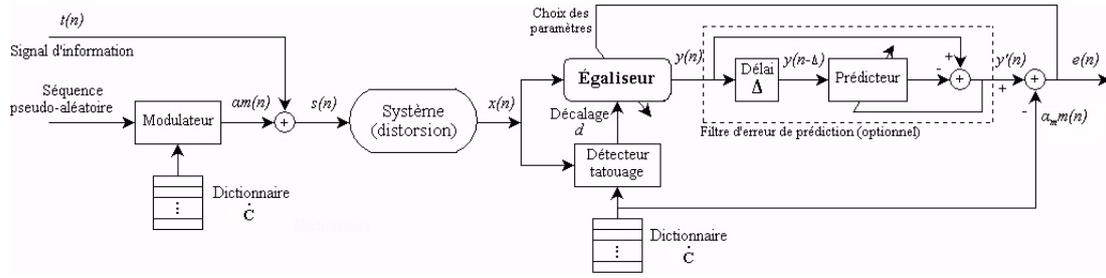


FIG. 1 – Schéma d'égalisation utilisant un tatouage comme signal de référence.

Pour trouver le décalage de la fenêtre glissant qui correspond le mieux à l'emplacement de chaque symbole, un algorithme de programmation dynamique [6] est utilisé. Le processus d'optimisation minimise une fonction coût exprimée par les matrices **A** et **B**. À l'aide d'un ensemble de J valeurs définies par le décalage λ , $[\hat{\lambda}_0 \dots \hat{\lambda}_{J-1}]$, on trouve le chemin dans les matrices **A** et **B** à partir duquel la séquence de symboles du tatouage peut être obtenue.

La fonction coût $c(\lambda, \lambda', j)$ pour passer d'un état $[\lambda', j-1]$ à l'état $[\lambda, j]$ est composée de trois termes :

$$c(\lambda, \lambda', j) = c_1(\lambda, \lambda', j) + c_2(\lambda, j) + c_3(\lambda, \lambda', j) \quad (10)$$

Le premier terme pénalise le passage du décalage λ de l'état $[\lambda', j-1]$ à l'état $[\lambda, j]$, puisque le décalage doit rester constant lorsque la séquence d'apprentissage est respectée. Ce terme diminue donc l'effet de pics de corrélation aléatoires entre le signal reçu et le tatouage. Il s'exprime par :

$$c_1(\lambda, \lambda', j) = \eta_j (\lambda - \lambda')^2 \quad (11)$$

où le terme quadratique impose une pénalité qui augmente rapidement avec la taille du saut de λ à λ' (i.e. l'on suppose que la probabilité qu'un segment soit enlevé du signal ou ajouté à ce dernier diminue avec la longueur de ce segment). Le facteur η_j est défini par :

$$\eta_j = \begin{cases} \eta_{j-1} + \kappa_1 & \text{pour } \lambda_j \neq \lambda_{j-1} \\ \max(\eta_{j-1} - \kappa_2, \eta_0) & \text{sinon} \end{cases} \quad (12)$$

où κ_1 et κ_2 sont des constantes positives (en général, $\kappa_1 > \kappa_2$), λ_j est le décalage correspondant à la colonne j dans le chemin, et η_0 est initié avec une valeur positive. Cette définition évite des chemins en zigzag, car η_j tend à augmenter dans cette situation.

Le deuxième terme dans la définition de la fonction coût est lié à la mesure de corrélation présente dans la matrice **A** :

$$c_2(\lambda, j) = \rho \left(1 - \frac{\beta_{\lambda, j}}{\max_{\tilde{\lambda}} \beta_{\tilde{\lambda}, j}} \right) \quad (13)$$

où ρ est une constante positive. L'expression entre parenthèses prend des valeurs entre 0 (lorsque le décalage λ correspond à la plus forte corrélation) et 1 (lorsque la corrélation pour le décalage λ est nulle). Cette définition pénalise des décalages λ correspondant à de faibles corrélations.

Le troisième terme, qui renforce l'observation de la séquence de symboles du tatouage, est défini par :

$$c_3(\lambda, \lambda', j) = \begin{cases} \psi (\chi_{\lambda, j} - \chi_{\lambda', j-1} - 1) & \text{pour } \chi_{\lambda, j} \geq \chi_{\lambda', j-1} \\ \psi (\chi_{\lambda, j} - \chi_{\lambda', j-1} - 1 + \hat{K}) & \text{sinon} \end{cases} \quad (14)$$

où ψ est une constante et \hat{K} le nombre de vecteurs dans le dictionnaire. Si la séquence de symboles du tatouage est respectée, ce coût est nul ; sinon, le coût est proportionnel à la largeur du saut observé. Cette définition se justifie par le fait que, en général, la probabilité qu'un segment de signal soit enlevé ou ajouté diminue avec la longueur de ce segment.

On définit le coût accumulé $C(\lambda, j)$ comme le coût minimal pour passer à l'état $[\lambda, j]$ à partir de la première colonne ($j = 0$). Ce coût est initialisé à 0 pour $j = 0$ et tout λ . L'algorithme d'optimisation est le suivant :

$$\begin{aligned} &\text{Pour } j = 1 \dots J - 1 \\ &\quad \text{Pour } \lambda = -\Lambda \dots \Lambda - 1 \\ &\quad \quad \lambda = \arg \min_{\lambda'} [C(\lambda', j-1) + c(\lambda, \lambda', j)] \\ &\quad \quad C(\lambda, j) = C(\tilde{\lambda}, j-1) + c(\lambda, \tilde{\lambda}, j) \\ &\quad \quad I(\lambda, j) = \tilde{\lambda} \\ &\hat{\lambda}_{J-1} = \arg \min_{\tilde{\lambda}} [C(\tilde{\lambda}, J-1)] \\ &\text{Pour } j = J - 2 \dots 0 \\ &\quad \hat{\lambda}_j = I(\lambda_{j+1}, j+1) \end{aligned}$$

Le résultat de l'algorithme est un ensemble de décalages $[\hat{\lambda}_0 \dots \hat{\lambda}_{J-1}]$ correspondant au chemin optimal.

4 Résultats

Pour illustrer la performance de l'algorithme, on présente sur la figure 2 l'erreur quadratique moyenne (EQM) du signal égalisé en fonction du rapport signal à tatouage (RST) pour des signaux numériques (non corrélé) et audio (corrélé). Le système de communication a été modélisé comme un filtre FIR de fonction de transfert $H(z) = 1 + 1.2z^{-1} - 0.3z^{-2} + 0.8z^{-3}$. L'algorithme RLS (*Recursive Least Squares*) a été utilisé avec un égaliseur à cinq coefficients. Pour un RST variant de 20 à 23 dB, l'EQM s'approche de sa valeur optimale de Wiener.

Pour illustrer le processus de synchronisation, 5 échantillons ont été enlevés aléatoirement du signal toutes les 200 fenêtres d'analyse. La longueur des fenêtres a été $N = 512$, avec un décalage maximal $\Lambda = 256$ pour la fenêtre glissante.

L'algorithme de programmation dynamique a utilisé les valeurs constantes $\psi = \Lambda$, $\eta_0 = 1$, $\kappa_1 = 5$, $\kappa_2 = 1$ et

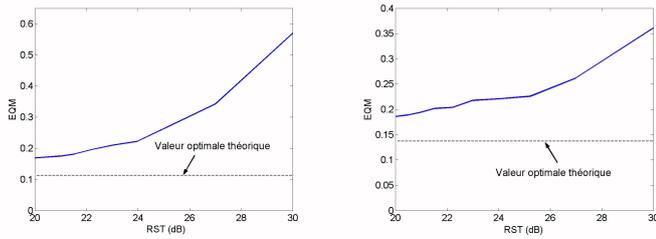


FIG. 2 – Performance de l’algorithme pour plusieurs valeurs de RST : signal numérique (à gauche) et signal audio (à droite).

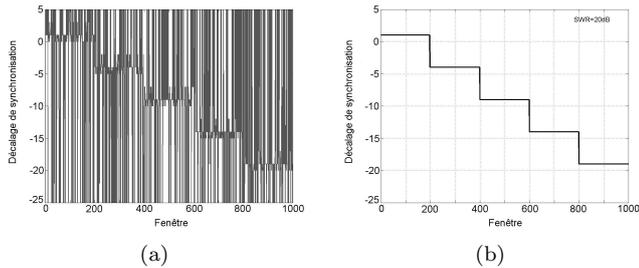


FIG. 3 – Performance de l’algorithme de synchronisation pour un signal numérique : 3(a) pics aléatoires de corrélation ; 3(b) programmation dynamique.

$\rho = 10$. Ces paramètres ont été ajustés empiriquement.

Les figures 3 et 4 présentent les résultats de cet algorithme pour les deux types de signaux. Les figures 3(a) et 4(a) illustrent les corrélations les plus fortes dans la matrice \mathbf{A} . On observe qu’il y a plusieurs pics aléatoires de corrélation, ce qui rend difficile la détermination du délai optimal de synchronisation à partir de ces valeurs. Les figures 3(b) et 4(b) montrent que l’algorithme de programmation dynamique a été capable de suivre la synchronisation de façon efficace, même lors des transitions.

5 Conclusions

La méthode d’égalisation adaptative utilisant un tatouage comme signal de référence présente des avantages importants par rapport aux techniques traditionnelles d’égalisation, notamment le fait que l’émission d’information utile n’est jamais interrompue. Elle s’applique aussi bien à l’égalisation de canaux de communication numérique

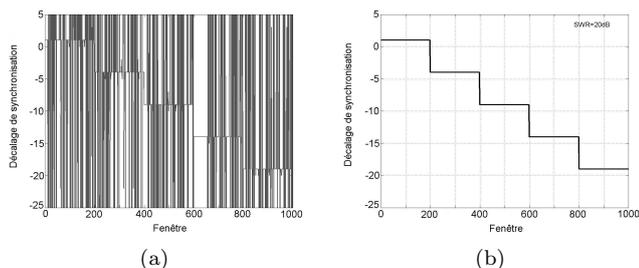


FIG. 4 – Performance de l’algorithme de synchronisation pour un signal audio : 4(a) pics aléatoires de corrélation ; 3(b) programmation dynamique.

qu’aux systèmes audio et de parole.

Une technique de synchronisation utilisant une séquence d’apprentissage présente dans le tatouage a été proposée. Cette technique permet de suivre continuellement la synchronisation du système.

Les résultats de simulation montrent que le système a réussi à égaliser le signal reçu et a été capable de retrouver rapidement la synchronisation lorsque des segments de signal ont été enlevés ou ajoutés.

Si la puissance du tatouage augmente, la détection du tatouage est plus précise et l’algorithme d’égalisation devient plus performant, ce qui mène à une meilleure détection du signal portant l’information utile. En revanche, lorsque la puissance du signal total émis est constante, l’augmentation de la puissance du tatouage entraîne une réduction correspondante de la puissance du signal portant l’information utile et, par conséquent, une détection moins efficace de ce dernier. En plus, la puissance du tatouage est limitée par la contrainte de transparence. Il y a donc un compromis entre la puissance du signal portant l’information utile et celle du tatouage pour que le système travaille dans des conditions optimales. Ce compromis sera étudié en détail dans nos recherches futures.

Dans les prochaines étapes de ce travail, on prévoit l’étude de la performance de la méthode devant des attaques plus sévères, telles qu’un filtrage non linéaire ou un processus de codage/décodage. Pour les applications qui concernent l’audio et la parole, on doit aussi considérer l’utilisation de modèles perceptifs pour augmenter la puissance du tatouage sans introduire de dégradation audible. La méthode sera aussi comparée avec des techniques classiques d’égalisation en termes de performance et de complexité.

Références

- [1] M. Uliani Neto, L. de C. T. Gomes, J. M. T. Romano et M. Bonnet. *Adaptive equalization based on watermarking*. VI International Telecommunications Symposium (ITS 2006), Fortaleza, Brésil, septembre 2006.
- [2] M. Uliani Neto. *Identificação e equalização em sistemas adaptativos utilizando marca d’água digital*. Thèse de maîtrise, FEEC/Unicamp, Campinas, Brésil, 2007.
- [3] J. K. Tugnait et W. Luo. *On channel estimation using superimposed training and first-order statistics*. IEEE Communications Letters, vol. 7, no. 9, septembre 2003.
- [4] L. de C. T. Gomes. *Tatouage de signaux audio*. Thèse de doctorat, Université Paris V, Paris, France, juillet 2002.
- [5] L. de C. T. Gomes, E. Gómez, M. Bonnet et N. Moreau. *Méthodes de resynchronisation pour le tatouage audio*. GRETSI’01, Toulouse, septembre 2001.
- [6] D. P. Bertsekas. *Dynamic programming and optimal control*. Athena Scientific, second edition, 2000.